

# nature

A detailed photograph of a blue butterfly with black spots and markings on its wings, perched on a pink flower. The background is a soft, out-of-focus yellowish-brown.

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

## FLORAL PATTERNS

Charting the links  
between flowers,  
nectar resources and  
insect pollinators

PAGE 85

### CLIMATE CHANGE

#### TROPICS ON THE MARCH

Tropical expansion threatens  
spreading drought

PAGE 20

### REPRODUCIBILITY

#### WRITING WRONGS

It's too hard to correct  
published errors

PAGE 27

### ELECTROCHEMISTRY

#### DIVIDING THE WATER

Designer catalysts for  
solar-energy conversion

PAGES 36 & 77

NATURE.COM/NATURE

4 February 2016 £10

Vol. 530, No. 7559



9 770028 083095

# THIS WEEK

## EDITORIALS

**EUROPE** Britain set to decide on continental drift **p.6**

**WORLD VIEW** Ben Goldacre on bad clinical-trial reporting **p.7**

**FLOODS** Climate-change simulations show heavier rain **p.8**



## The next steps on Zika

*With birth defects blamed on the virus now deemed a matter of international concern, researchers must work fast to assess the extent of the threat.*

**T**he World Health Organization this week declared that clusters of birth defects suspected of being linked to an epidemic of Zika virus in the Americas constituted a “public health emergency of international concern”. Beyond the practical imperative to better control the mosquitoes that spread Zika and other diseases, the most urgent priority on the ground is research to answer basic, but crucial, questions, including whether the birth defects are caused by the virus, and if so, how frequently.

Alarm among media and politicians that the Zika virus is poised to cause a pandemic of birth defects has reached fever pitch — with some headlines even screaming that it could be worse than Ebola in West Africa. But, as in the early stages of any epidemic of a new or re-emerging virus, there are challenges in dealing with the uncertainties that inevitably arise from huge gaps in our knowledge.

A calm and cautious approach to assessing the threat is advisable, and so is avoiding unhelpful hype and hysteria, or jumping to premature conclusions. Health authorities must, of course, not wait until all the facts are in before taking action, and are right to recommend, out of an abundance of caution, measured precautions to protect pregnant women and their fetuses in the face of a perceived potential threat.

The Zika virus is transmitted mainly by the *Aedes aegypti* mosquito, which also carries the dengue and chikungunya viruses. It is unusual in that it has evolved to breed in urban areas in small pockets of water such as in discarded tyres and pot-plant trays. It feeds exclusively on human blood, and during the day, making it a formidable disease threat across huge swathes of tropical and subtropical regions. To a lesser extent, at least so far, Zika can also be spread by *Aedes albopictus*, whose distribution reaches far into temperate zones.

First discovered in a monkey in the Zika forest of Uganda in 1947, the virus has until recently been little more than a scientific curiosity. Around 80% of those infected seem to show no symptoms, and the remainder have no more than a mild illness that gets better by itself in a few days. Originally restricted to small outbreaks in Africa and Asia, the virus fanned out to the Pacific in 2007, and since last year has — like dengue and chikungunya before it — spread rapidly through the tropical and subtropical regions of the Americas.

### MEASURE OF CONTROL

Concerns that this mild clinical picture may not be the whole story first arose last October, when doctors in Brazil reported that a large epidemic of Zika seemed to be associated with a spike in the number of cases of microcephaly, a serious congenital malformation in which fetuses and newborns have unusually small heads and brains. Brazil officially reported just 147 cases of microcephaly in 2014; since October it has recorded more than 4,000 suspected cases.

The apparent steep rise is terrifying, especially for pregnant women and their families. But is it real? The 147 cases reported in Brazil

in 2014 are surely a vast underestimate. Even using conservative estimates, the annual frequency of microcephaly in the country should be around an order of magnitude higher.

Moreover, since October, Brazilian authorities have launched an intensive search for cases of microcephaly. As we report on page 13, that scrutiny will turn up many cases that would otherwise have gone undetected, and the broad screening criteria used mean that many of the suspected cases will inevitably turn out to be false alarms.

At this stage, we simply do not have enough epidemiological and clinical data to fully assess how much of a threat Zika poses. Gathering this information must be the priority. It is incumbent on researchers and authorities in Brazil and other countries affected by Zika to put aside competition for papers, or concerns about making data

**“Zika virus yet again highlights the problem of how best to prepare against emerging diseases that may — or may not — pose a threat.”**

public, and publish all available epidemiological and clinical data as soon as they are gathered and analysed.

It will be particularly crucial to establish the scale of any increase in birth defects that is attributable to Zika. It would be something of a relief, for example, if Zika was quickly found to be unlike rubella (German measles), which infected millions of people and caused tens of thousands of birth defects in the 1960s in the United States alone. Even though a cheap and effective vaccine has now largely wiped out rubella in richer countries, the disease still causes birth defects in some 100,000 babies worldwide every year.

Until better data become available on Zika, we should hope for the best, but prepare for the worst. Circumstantial evidence is slowly accumulating that the virus may be associated with birth defects; it has been found in the amniotic fluid, placenta or fetal tissue of a small number of those affected. Also, although hard data are crucial, so too are the first-hand accounts and perceptions of clinicians who treat babies with microcephaly. They might be subject to some confirmation bias, but the clinicians are reporting a clear signal in Brazil that something could be amiss.

More broadly, the Zika virus yet again highlights the problem of how best to prepare against emerging diseases that may — or may not — pose a serious threat to public health. Gaining ground is the sensible idea that it would be a relatively inexpensive insurance policy to develop experimental treatments, and above all vaccines, against an array of potential threats.

The most important immediate step is to attack the *Aedes* mosquito vector. Effective vector control is not easy, but an all-out campaign is in any case long overdue. Even if the threat of Zika turns out to have been overrated, concerted action now would have valuable pay-offs in reducing the burden of the established disease threats of dengue and chikungunya in the Americas. ■



# Green growth

*US policymakers must set aside their divisions and give climate research a much-needed boost.*

Columbia University in New York has invested US\$3.1 million in a new centre to bolster climate research while building ties with big business and corporate philanthropists. The idea is simple: the business community has both the resources to promote climate science and a vested interest in the results. The university, meanwhile, is looking for a stable funding source to support its research. Other institutions have taken similar paths, and as long as the science remains independent, everybody benefits. But even if the Center for Climate and Life succeeds, it does not solve the larger problem of stagnant US budgets for environmental research.

The proposal for the centre initially arose from Columbia's Lamont-Doherty Earth Observatory in Palisades, which has nearly 200 Earth scientists. Most are dependent on research grants for the majority of their salaries. As a News story on page 16 makes clear, it is particularly difficult for young researchers there to get government funding. But even the senior scientists are coming up empty-handed with increasing frequency, and that means more time spent on grant applications and less time spent on science. Money is not always the answer, but federal funding has not kept up with the times.

The numbers say it all. Adjusted for inflation, the US budget for environmental science peaked in fiscal year 2003 at \$4.73 billion and then began to decline. Things have improved over the past few years, but only slightly. The budget for environmental science stood at \$4.18 billion in 2015, the latest year for which data are available. That is roughly the same as more than two decades ago, but it is safe

to say that awareness of the environmental threats posed by resource consumption, rising populations and global warming has increased significantly in that time.

One factor was the financial crisis of 2008. It has taken time for the United States and other countries to recover; there has simply not been as much money to go round. But no comfort has come from the US Congress, where conservatives have sought — unsuccessfully so far — to further reduce spending on climate sciences. Long-standing partisan divides over climate and environmental policy have begun to seep into the debates over funding for basic science, which in the past has remained relatively bipartisan.

In 2007, the US Congress passed the America COMPETES Act, which enabled a doubling of science budgets at several agencies over ten years. America COMPETES was re-authorized in 2010, but lawmakers have yet to follow through with increased funding in the annual appropriations process. Last year, Republicans in the House of Representatives passed yet another re-authorization of the act, but this controversial version would have actually reduced research spending in various areas, including a 12% reduction in funding for geosciences. It drew a veto threat from US President Barack Obama, and rightly so.

The original drive for America COMPETES followed an influential 2005 report by the National Academies of Science, Engineering, and Medicine, titled *Rising Above the Gathering Storm*. It argued that US economic competitiveness depends on innovation in energy, science and technology, and this remains true today. Policymakers on both sides of the aisle say that they want the United States to lead. They say that they want decisions to be based on solid science. But leadership requires energy, and science requires investment — as well as freedom from political interference. ■

**“Partisan divides have begun to seep into the debates over funding for basic science.”**

# Better together

*The European Union has its issues, but a Brexit could spell problems for science.*

Greece narrowly avoided being kicked out of the European Union's single currency last year. Now Britain could exit the whole union voluntarily (a possibility dubbed 'Brexit'). The grand EU project has not looked so shaky since the financial crash of 2008, which brought many of the 28 EU member states to their knees. Now, in the midst of a Europe-wide migration crisis, the United Kingdom is trying to renegotiate its EU membership. If concessions won this week do not convince the UK public, then Britain could vote to leave the EU by the end of 2017.

Many researchers in the United Kingdom and elsewhere are aghast at this idea. As detailed on page 15, UK scientists pull in millions of euros of funding from EU research programmes and gain preferential access to major infrastructure projects. No one is really sure whether Britain leaving would jeopardize this, because no state has left the EU before.

British discussion of the relationship with Europe is coloured by decades of tabloid scorn and political opportunism. The first step to any sensible decision must be to separate the facts from the rhetorical fiction. So it is essential that debates over the role of the EU in science — such as the ongoing inquiry by the House of Lords science committee into the EU's influence — offer a critical analysis of what science stands to lose.

It is certainly true that EU funding has sustained many a scientific career as national-government support has withered throughout the continent. Across leading British universities, one in every five employees comes from a non-UK EU member state. UK scientists ply their trade

in universities across the EU in return. Indeed, for many scientists, the country that they work in is less important than the work that they do.

Opponents of a Brexit make a compelling case for what might be lost if UK researchers were locked out of EU systems and UK institutions had to jump through hoops to bring in talented staff. Among those who have come out to sing the praises of EU links are the academic umbrella group Universities UK, the elite-university lobbying organization the Russell Group, and the anti-Brexit group 'Scientists for EU'.

They argue that cooperation in research shows Europe at its best. British citizens were as proud as any across the continent, for example, when the European Space Agency wowed the world last year by landing a spacecraft on a comet. And the CERN particle-physics laboratory is a model of trans-boundary science — quite literally, because its accelerators near Geneva straddle the border between France and Switzerland.

There have been complaints that debate might be stifled by the great and the good in science coming out wearing their institutional badges and arguing against a Brexit. But scientists — to their credit — generally do not shy away from speaking up. A dedicated group of pro-Brexit researchers is making its own case.

The group's position has some merit, too. It would be wrong to pretend that the EU is without problems. Supporters of a Brexit cite rules on clinical trials as an example of EU regulations that have hindered the pursuit of knowledge. And the political gridlock on growing genetically modified crops in the EU has left agricultural scientists frustrated. Brussels bureaucracy has become a shorthand term for red tape, officialdom and delays.

On balance, it is the view of this journal that science, in Britain and elsewhere, would benefit from the United Kingdom remaining as a committed member of the European team. An exit decision would cause chaos and uncertainty, and could set back some projects significantly. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunqv](http://go.nature.com/xhunqv)



## Make journals report clinical trials properly

*There is no excuse for the shoddy practice of allowing researchers to change outcomes and goals without saying so, says Ben Goldacre.*

Science is in flux. The basics of a rigorous scientific method were worked out many years ago, but there is now growing concern about systematic structural flaws that undermine the integrity of published data: selective publication, inadequate descriptions of study methods that block efforts at replication, and data dredging through undisclosed use of multiple analytical strategies. Problems such as these undermine the integrity of published data and increase the risk of exaggerated or even false-positive findings, leading collectively to the ‘replication crisis’.

Alongside academic papers that document the prevalence of these problems, we have seen a growth in ‘technical activism’: groups creating data structures and services to help find solutions. These include the Reproducibility Project, which shares out the work of replicating hundreds of published papers in psychology, and Registered Reports, in which researchers can specify their methods and analytical strategy before they begin a study.

These initiatives can generate conflict, because they set out to hold individuals to account. Most researchers maintain a public pose that science is about healthy, reciprocal, critical appraisal. But when you replicate someone’s methods and find discrepant results, there is inevitably a risk of friction.

Our team in the Centre for Evidence-Based Medicine at the University of Oxford, UK, is now facing the same challenge. We are targeting the problem of selective outcome reporting in clinical trials.

At the outset, those conducting clinical trials are supposed to publicly declare what measurements they will take to assess the relative benefits of the treatments being compared. This is long-standing best practice, because an outcome such as ‘cardiovascular health’ could be measured in many ways. So researchers are expected to list the specific blood tests and symptom-rating scales that they will use, for example, alongside the dates on which measurements will be taken, and any cut-off values they will apply to turn continuous data into categorical variables.

This is all done to prevent researchers from ‘data-dredging’ their results. If researchers switch from these pre-specified outcomes, without explaining that they have done so, then they break the assumptions of their statistical tests. That carries a significant risk of exaggerating findings, or simply getting them wrong, and this in turn helps to explain why so many trial results eventually turn out to be incorrect.

You might think that this problem is so obvious that it would already be competently managed by researchers and journals. But that is not the case. Repeatedly, academic papers have been published showing that outcome-switching is highly prevalent, and that such switches often lead to more favourable statistically significant results being reported instead. This is despite

numerous codes of conduct set up to prevent such switching, most notably the widely respected CONSORT guidelines, which require reporting of all pre-specified outcomes and an explanation for any changes. Almost all major medical journals supposedly endorse these guidelines, and yet we know that undisclosed outcome-switching persists.

Our group has taken a new approach to trying to fix this problem. Since last October, we have been checking the outcomes reported in every trial published in five top medical journals against the pre-specified outcomes from the registry entries or protocols. Most had discrepancies, many of them major. Then, crucially, we have submitted a correction letter, on every trial that misreported its outcomes, to the journal in question. (All of our raw data, methods and correspondence with journals are available on our website at [COMParE-trials.org](http://COMParE-trials.org).)

We expected that journals would take these discrepancies seriously, because trial results are used by physicians, researchers and patients to make informed decisions about treatments. Instead, we have seen a wide range of reactions. Some have demonstrated best practice: the *BMJ*, for instance, quickly published a correction on one misreported trial we found, within days of our letter being posted.

Other journals have not followed the *BMJ*’s lead. The editors at *Annals of Internal Medicine*, for example, have responded to our correction letters with an unsigned rebuttal that, in our view,

raises serious questions about their commitment to managing outcome-switching. For example, they repeatedly (but confusedly) argue that it is acceptable to identify ‘prespecified outcomes’ from documents produced after a trial began; they make concerning comments that undermine the crucial resource of trial registers; and they say that their expertise allows them to permit — and even solicit — undeclared outcome-switching. Furthermore, they have declined to publish our response to their 850-word letter in the journal.

In our view, this is troubling. *Annals*’ response helps to explain why studies repeatedly find outcome-switching to be hugely prevalent, despite policies to prevent it. But journal editors now need to engage in a serious public discussion on why this is still happening. We are providing specific worked examples to facilitate this discussion, and if our project is regarded as provocative, then that is misguided. Audit and accountability are the bread and butter of good medicine, and good science. Lives are at stake when subtle statistical signals of benefit and risk are sought in noisy, messy trial data. We hope that the structures of science really are in a state of flux, and still changing. ■

**Ben Goldacre** is a physician, author and senior clinical research fellow at the University of Oxford, UK.  
e-mail: [ben.goldacre@phc.ox.ac.uk](mailto:ben.goldacre@phc.ox.ac.uk)

**AUDIT AND  
ACCOUNTABILITY  
ARE THE BREAD AND  
BUTTER OF GOOD  
MEDICINE, AND  
GOOD SCIENCE.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/8wdqhd](http://go.nature.com/8wdqhd)



# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## ASTROBIOLOGY

### Fungus survives simulated Mars

A small percentage of fungal cells were still able to divide after exposure to Mars-like conditions aboard the International Space Station.

Dried samples of the Antarctic-dwelling black fungi *Cryomyces antarcticus* and *Cryomyces minteri*, which live inside rocks, were exposed for 18 months to a simulated Martian atmosphere of 95% carbon dioxide, as well as high levels of ultraviolet and cosmic radiation. Silvano Onofri at the University of Tuscia in Viterbo, Italy, and his colleagues found that less than 10% of the samples divided and formed colonies after their return to Earth. However, up to two-thirds of the cells remained intact and yielded stable DNA.

The findings could inform future searches for evidence of life on Mars, the authors say. *Astrobiology* 15, 1052–1059 (2016)

## ANIMAL BEHAVIOUR

### Colour signals octopus fight

Octopuses use body colour and posture to communicate to others during aggressive encounters, suggesting that they are more social than previously thought.

Octopuses are considered to be more solitary animals than many squid or cuttlefish. David Scheel of Alaska Pacific University in Anchorage and his colleagues reviewed filmed interactions between pairs of *Octopus tetricus* off the coast of New South Wales, Australia. Octopuses were darker when they were about to fight an approaching animal and paler when they were set to flee. When dark in colour, the animals also changed their

posture — by standing tall on higher ground and spreading the webs between their arms.

The octopuses could be using these signals to communicate their size, strength and willingness to fight to a rival, the authors say. *Curr. Biol.* <http://doi.org/bb6f> (2016)

## METABOLISM

### On-off switch for obesity

Genetically identical mice either become obese or stay lean because of differences in how a specific set of genes is regulated.

J. Andrew Pospisilik at the Max Planck Institute

of Immunobiology and Epigenetics in Freiburg, Germany, and his colleagues studied mice with a mutation in the protein TRIM28, which has been linked to variation in body mass. The mice fell into two groups: those that grew obese and those of a normal weight, with the obese animals showing reduced expression of a network of genes involved in controlling weight.

The researchers then looked at data from obese and normal-weight children, as well as from 13 pairs of identical twins — in which one twin was obese and the other was lean — and found similar patterns of gene expression to the TRIM28 mutant mice. Differences in maternal diet, communities

past and present climate conditions. In the model runs that were based on current conditions, the likelihood of extremely wet winters was about 43% higher than in simulations with pre-industrial concentrations of greenhouse gases.

Climate warming increased moisture and shifted circulation patterns in the atmosphere, boosting the risk of heavy precipitation.

*Nature Clim. Change* <http://dx.doi.org/10.1038/nclimate2927> (2016)

## CLIMATE SCIENCE

### UK floods tied to climate change

The risk of extreme winter rainstorms in southern England, such as those that caused widespread flooding in January 2014, has increased as a result of climate warming.

Harnessing spare processing time on volunteers' personal computers, Nathalie Schaller and Neil Massey at the University of Oxford, UK, and their colleagues produced more than 134,000 simulations of possible winter weather in southern England and Wales under



FACUNDO ARRIZABALAGA/EPA/CORBIS

## CONSERVATION

### A map of threats from alien species

Australia, the Americas and India are among the global hotspots where animals are most threatened by invasive species.

Ecologists debate the extent to which biodiversity is threatened by non-native species compared to threats from land-use change and other factors. To explore this, Céline Bellard at University College

London and her colleagues analysed the ranges of nearly 1,400 threatened vertebrate species and more than 200 non-native organisms worldwide. The authors also used two databases that list the causes of the animals' decline. They found that 22% of all threatened animals are imperilled by invasive organisms, but none was threatened by non-natives alone.

The invasive organism that threatened the most vertebrates was a lineage of the chytrid fungus *Batrachochytrium dendrobatidis*, which has killed many amphibians.

*Proc. R. Soc. B* 283, 20152454 (2016)

## MICROBIOLOGY

## What makes *Listeria* deadly

A group of six genes causes some strains of a foodborne bacterium to become highly dangerous, or virulent.

*Listeria monocytogenes* (pictured) can be found in many foods, including unpasteurized milk, and can cause miscarriage in pregnant women or kill infected people if the pathogen moves to the brain. All strains are currently considered to be equally virulent, but Sylvain Brisse and Marc Lecuit of the Pasteur Institute in Paris and their colleagues found differences in virulence when they analysed genomic and epidemiological data from more than 6,600 *L. monocytogenes* isolates taken from food and human samples. Food-associated strains mainly infected people who had weakened immune systems, but more-virulent clinical isolates were found in healthy people. The 'CC4'

group of strains carried a six-gene cluster that, when deleted, made the bacteria less capable of invading brain and placental tissues in mice.

Public-health surveillance efforts for foodborne illnesses should look out for these hypervirulent strains, the authors say.

*Nature Genet.* <http://dx.doi.org/10.1038/ng.3501> (2016)

## ASTRONOMY

## Widest-known planetary orbit

A planet takes 900,000 years to orbit its star, with the two objects separated by a distance almost 7,000 times that between Earth and the Sun.

Niall Deacon at the University of Hertfordshire in Hatfield, UK, and his colleagues found that the previously known star and the planet — located about 32 parsecs (104 light years) from Earth — move across the sky at the same speed and in the same direction, which suggests that the two form a pair.

The star, TYC 9486-927-1, is young, at just a few tens of millions of years old. And the planet has a mass between 12 and 15 times that of Jupiter — which makes it big and bright enough to be spotted from Earth through a telescope, unlike most other exoplanets.

*Mon. Not. R. Astron. Soc.* in the press; preprint at <http://arxiv.org/abs/1601.06162> (2016)

## GEOPHYSICS

## Unnatural shaking in California

Oil and gas exploration could be triggering earthquakes in California, where natural quakes might mask the induced ones.

The injection of millions of litres of wastewater into underground oil and gas reserves has been linked to earthquakes around the world, including in central parts of the United States, such as Oklahoma. Thomas Goebel of the University of



California, Santa Cruz, and his colleagues analysed three quakes of magnitude-4 and higher that happened on the same day in September 2005 along southern California's White Wolf fault. They found unusual seismic patterns and, using a hydrogeological model, concluded that nearby fluid injection that began just months before was probably responsible.

More analysis is needed to identify human-induced earthquakes in California, the authors say.

*Geophys. Res. Lett.* <http://doi.org/bb6k> (2016)

## ECOLOGY

## Bat cave choice affects disease

The susceptibility of bats to a deadly disease is influenced by the humidity of the caves in which they hibernate.

White-nose syndrome, which has killed millions of bats in North America, is caused by infection with the fungus *Pseudogymnoascus destructans*. The disease is thought to speed up bats' use of energy reserves during hibernation, putting the animals at risk of starvation.

To find out why only some bat populations survive these infections, David Hayman at Massey University in Palmerston North, New Zealand, and his team used a computer model to predict winter survival times for four species. Humidity significantly reduced survival, which could explain why the little brown bat (*Myotis lucifugus*; pictured) — a resident of warm, humid

caves during hibernation — has been affected more than other bat species.

Larger bats, including some that live in Europe, are more likely to survive the infection. *Sci. Adv.* 2, e1500831 (2016)

## MICROBIAL ECOLOGY

## Archaea revealed from genomes

Researchers have pieced together the genomes of a newly identified group of single-celled organisms, unearthing clues about their metabolism.

Archaea are different from bacteria and are generally less studied. To find new, uncultured archaea, Brett Baker at the University of Texas at Austin in Port Aransas and his colleagues extracted and analysed DNA in underwater sediment cores from estuaries in North Carolina. They reconstructed three near-complete genomes of archaea belonging to a new group that the team called Thorarchaeota. The organisms have many genes for breaking down proteins and transporting amino acids, suggesting that proteins are their main carbon source. The microbes could also be biochemically cycling sulfur in sediments.

Identification of these organisms fills in a key part of the archaeal tree of life, the authors say.

*ISME J.* <http://doi.org/bb7p> (2016)

► NATURE.COM

For the latest research published by Nature visit:

[www.nature.com/latestresearch](http://www.nature.com/latestresearch)





# SEVEN DAYS

The news in brief

## EVENTS

### Zika emergency

The World Health Organization (WHO) has declared a global public-health emergency over the cluster of birth-defect cases linked to the outbreak of Zika virus in Brazil and the Americas. The mosquito-borne virus has been linked to a condition called microcephaly, in which babies are born with abnormally small heads and brains, in children born to infected mothers. WHO experts say that the link is strongly suspected, if not yet scientifically proven. According to the Pan American Health Organisation, the disease is now present in 26 countries and territories in the Americas. The WHO says that restrictions on travel and trade are not necessary. See pages 5 and 13 for more.

### Scandal memoirs

Haruko Obokata, the researcher at the centre of a stem-cell scandal, released a memoir on 28 January. Obokata's claims that embryonic-like stem cells could be created by stressing mature mouse cells in a process called stimulus-triggered acquisition of pluripotency (STAP) fell apart, and two *Nature* papers were retracted in July 2014. In the book, Obokata attempts to shift responsibility for part of the scandal to a former mentor. Fifty thousand copies of the book were released initially. According to publisher Kodansha, many stores have sold out.

## RESEARCH

### Embryo editing

In a world first, the UK Human Fertilisation and Embryology Authority has given scientists permission to edit the genomes of human embryos for research. The decision, announced

on 1 February, allows developmental biologist Kathy Niakan at the Francis Crick Institute in London to use the genome-editing technique CRISPR-Cas9 in healthy human embryos. Niakan's team plans to alter genes that are active in the first few days after fertilization. After seven days, the experiments will end and the embryos will be destroyed. See page 18 for more.

### Ancestry error

Software problems have forced researchers to go back on claims that humans across the whole of Africa carry DNA inherited from Eurasian immigrants. The authors issued a note explaining the mistake in an October 2015 *Science* paper on the genome

of a 4,500-year-old man from Ethiopia (M. Gallego Llorente *et al. Science* **350**, 820–822; 2015). The paper had confirmed findings that although the first humans began to leave Africa 100,000 years ago, a group returned some 3,000 years ago. But its conclusion that the backflow reached West and Central Africa was wrong. The authors say that they used incompatible software packages to compare genetic variants, causing the error. See [go.nature.com/tzfing](http://go.nature.com/tzfing) for more.

### Uranium glut

The world's civilian research reactors are taking too long to eliminate the use of highly enriched uranium, says a 28 January report from the

with the depth achieved. They intend to return and try to deepen the hole to reach the boundary between Earth's crust and mantle. The two-month expedition was sidetracked in part by a team member requiring medical evacuation to Mauritius in the middle of drilling.



XINHUA/ALAMY

## Ocean-floor drilling mission digs deep

An expedition in the Indian Ocean that aimed to create the deepest hole ever drilled into the ocean crust ended on 27 January, having reached 789 metres below the sea floor. Project scientists had hoped that drilling with the *JOIDES Resolution* ship would reach at least 1.3 kilometres, but say they are happy

US National Academies of Science, Engineering, and Medicine. Despite calls to phase out weapons-grade uranium, which is enriched to 90% or greater in the uranium-235 isotope, 74 reactors continue to use it. As an interim step to phasing it out completely — a process likely to take two decades — the report recommends blending the fuel to 45% enrichment or less.

## BUSINESS

### Diagnostics deal

Global health-care company Abbott is to buy diagnostics firm Alere for US\$5.8 billion. The buyout, announced on 1 February, will bolster Abbott's diagnostics portfolio. Alere, in Waltham, Massachusetts, has

annual sales of \$2.5 billion and makes a range of point-of-care tests that can be used in remote locations to detect infections including HIV, tuberculosis, malaria and dengue. If the deal is approved, the firm will become an Abbott subsidiary, and Alere shareholders will get \$56 per share.

## POLICY

**EU car tests**

The European Commission has unveiled tougher vehicle standards for car manufacturers in the wake of the Volkswagen emissions scandal. The proposed rules, announced on 27 January, will include increased independence of vehicle-emissions testing, spot checks on cars already on the market, and more powers for European authorities in an area that has been mainly the responsibility of individual governments. Volkswagen is still dealing with the aftermath of the revelation that it fitted some vehicles with devices that were designed to fool emissions tests. The company faces legal action from governments and consumers.

**Fish flip-flop**

Despite the decision last November by the US Food and Drug Administration (FDA) to approve genetically modified salmon for human



consumption, Americans will not be eating the fish any time soon. On 29 January, the FDA banned imports of fast-growing salmon produced in Panama and Canada by AquaBounty Technologies of Maynard, Massachusetts. The move is in response to the US budget bill passed last December, which bans sales of the fish until the FDA decides whether it should be labelled as genetically modified. The agency may take several years to finalize this rule.

**Pipeline policy**

Pipelines and other natural-resource projects in Canada will soon be subject to more-robust environmental and public-review processes. The new assessment, announced on 27 January, will analyse greenhouse-gas emissions from construction and processing, and will include a broader public-consultation process, emphasizing the rights of indigenous peoples.

The announcement is part of a broader effort by Prime Minister Justin Trudeau to bolster action on climate change. Last November, the United States denied Canadian energy-infrastructure company TransCanada a construction permit for a major pipeline that would have carried oil from the tar sands of Alberta to US refineries.

## PEOPLE

**Misconduct penalty**

Plant biologist Olivier Voinnet of the Swiss Federal Institute of Technology in Zurich has been stripped of an award and given a three-year ban on receiving funds from the Swiss National Science Foundation. Voinnet made his name studying gene silencing in plants, but charges of image manipulation and other issues have led to a slew of retractions and corrections of his published work (see *Nature* <http://doi.org/bb7z>; 2015). On 28 January, Europe's

## COMING UP

**5 FEBRUARY**

The evolutionary origins of the genus *Homo* are debated at the Salk institute in La Jolla, California, in a seminar hosted by the Center for Academic Research and Training in Anthropogeny.  
[go.nature.com/alzjp7](http://go.nature.com/alzjp7)

**7–11 FEBRUARY**

Researchers looking at the cancer genome gather in Banff, Canada.  
[go.nature.com/abjawy](http://go.nature.com/abjawy)

life-sciences organization EMBO told its members that Voinnet had agreed to return the EMBO Gold Medal he won in 2009.

**Trachea film**

A documentary has rekindled controversy surrounding surgeon Paolo Macchiarini, who is famed for performing artificial-trachea transplants. Concerns raised in the documentary, aired last month by SVT, Sweden's public broadcaster, have prompted the Karolinska Institute in Stockholm, where Macchiarini is a visiting professor, to investigate — five months after it cleared him of separate accusations of scientific misconduct. Footage in the film suggests that one patient who received the risky procedure and later died was in broadly good health before the operation, contradicting the surgeon's statements that he attempted the surgery only on people facing life-threatening illness. The documentary also illuminated a mismatch between the actual performance of a prosthetic trachea and its description in a paper co-authored by Macchiarini. See [go.nature.com/ecqve3](http://go.nature.com/ecqve3) for more.

**NATURE.COM**

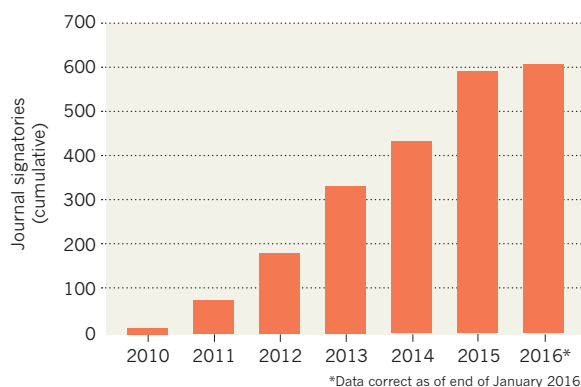
For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

More than 600 journals have signed up to the ARRIVE guidelines, which are designed to improve the reporting of animal experiments. Released in 2010, they list details that should be in such reports, including strains and sexes of animals, calculations and adverse reactions. More journals than ever signed up in 2015, says the UK National Centre for the Replacement, Refinement and Reduction of Animals in Research, which produced the guidelines. Some have questioned whether ARRIVE can be enforced.

**SURGE IN SUPPORT FOR STUDY GUIDELINES**

In 2015, more than 150 journals signed up to the ARRIVE checklist for animal studies — the highest number of signatories in a single year since it was released.





# NEWS IN FOCUS

**POLITICS** Britain's EU vote could impact science across the region **p.15**

**PSYCHIATRY** Can drugs tackle depression's cognitive symptoms? **p.17**

**GERMANY** Science clusters are clear winners of funding revamp **p.18**

**ROBOTICS** Soft, squishy machines are in — hard, clunky ones out **p.24**



UESLEI MARCELINO/REUTERS



Mothers of children with suspected microcephaly await medical care at a hospital in Recife, Brazil.

## ZIKA VIRUS

# Microcephaly surge in doubt

*Heightened awareness of Zika virus could help to explain the reported spike in birth defects.*

BY DECLAN BUTLER

**A**s alarm grows about the rapid spread of the Zika virus through the Americas, researchers are questioning the extent of an apparent surge in Brazil of microcephaly — a birth defect in which babies are born with abnormally small heads and brains — which has been tentatively linked to the virus.

On 1 February, the possibility that microcephaly is linked to Zika prompted the World Health Organization (WHO) to label the situation a “public health emergency of

international concern”, a mechanism that will allow it to coordinate an international response, among other things.

But Jorge Lopez-Camelo and Ieda Maria Orioli, from the Latin American Collaborative Study of Congenital Malformations (ECLAMC), say that a rise in reported cases of microcephaly might largely be attributable to the intense search for cases of the birth defect and to misdiagnoses, arising from heightened awareness in the wake of the possible link with Zika.

The ‘awareness’ effect is well known and

inevitable, they say, and must be revealing cases that would have gone unnoticed under normal circumstances. And a high rate of misdiagnoses among reported cases of microcephaly is likely because the diagnostic criteria being used are broad. Lopez-Camelo and Orioli posted their analysis on the ECLAMC website in Portuguese, and, after *Nature*'s enquiries, provided an English-language version of the summary (see [go.nature.com/v77ntu](http://go.nature.com/v77ntu)).

They say that from the epidemiological data available, it is impossible to establish the ►

► true size of the surge in microcephaly, and whether there is any link with the Zika virus.

In particular, large prospective studies are needed, in which pregnant women in areas of Brazil experiencing Zika outbreaks would be monitored to see how many of their children develop microcephaly, say Lopez-Camelo and Orioli. Several research groups in and outside Brazil are already planning such studies, and some have begun.

Specialists contacted by *Nature* emphasize that pregnant women should still be cautious — for example, by protecting themselves against mosquito bites — until more is known.

These experts agree with the ECLAMC team that the reported size of the microcephaly increase so far is probably inflated — but some disagree with the further conclusion that the reported rise is mostly an artefact.

Thomas Jaenisch, a tropical-medicine specialist at the Heidelberg University Hospital in Germany, calls the attribution of most of the surge to these awareness effects an “extreme” position, and says that it “might also create uncertainty in the media and public discussion in Brazil”.

On 27 January, the Brazilian government said that of 4,180 suspected microcephaly cases recorded since October, it has so far confirmed 270 and rejected 462 as false diagnoses.

Previously confined to Africa and Asia, Zika virus reached the Americas in 2015, where it is now causing an unprecedented epidemic in Brazil and outbreaks in several countries (see ‘Zika in the Americas’). Most people infected with the virus — after being bitten by a mosquito — have no symptoms. The remainder have mild symptoms such as fever, skin rash and headache.

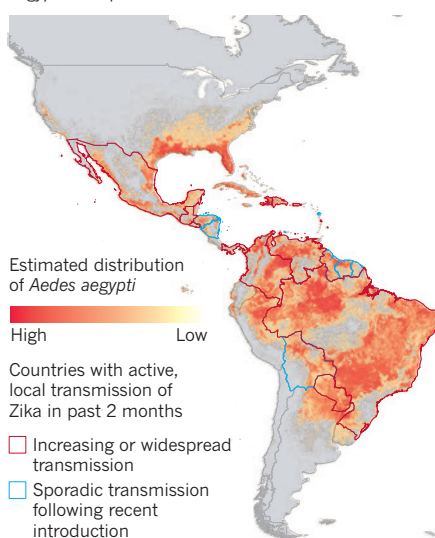
But in October, Brazil’s health ministry reported an unusual spike in reported cases of microcephaly in the northeastern state of Pernambuco, where the affected children’s mothers had been in early pregnancy at around the same time as large Zika outbreaks occurred. The ministry subsequently raised the alarm of a possible link to Zika.

This led the WHO and its regional office, the Pan American Health Organization, to issue an epidemiological alert on 17 November, which called on member states to look for any similar increase in microcephaly among their populations.

To investigate, the researchers at the

## ZIKA IN THE AMERICAS

Following its arrival in the Americas in 2015, Zika virus is now being actively transmitted in many of the countries that harbour its main carrier, the *Aedes aegypti* mosquito.



ECLAMC turned to its own databases dating back to 1967, as well as the country’s Live Birth Information System (SINASC).

According to the ECLAMC, the average historical prevalence of microcephaly in Brazil is around 2 cases per 10,000 births, although rates in the country’s north have typically been higher. The researchers calculate that the maximum number of cases that would have been expected in the northern state of Pernambuco in 2015 is around 45. Yet Pernambuco reported 26 times that number last year. Even if Zika is causing microcephaly, these huge numbers are simply too high to be credible, says the report.

## INCREASED SURVEILLANCE

As well as the increased diagnoses resulting from heightened awareness, Lopez-Camelo also highlights that the diagnostic criteria for microcephaly are relatively unspecific and cast too wide a net. Brazilian health authorities are treating all fetuses with head circumferences that are more than two standard deviations below the average, and newborns with a head circumference of less than 32 centimetres, as suspected cases. But these criteria will inevitably capture many healthy children within the normal growth range who do not have

microcephaly. And head circumference is only a proxy measure, note Lopez-Camelo and Orioli: confirming microcephaly requires a diagnosis of small brain size, and a decreased rate of brain growth.

A risk assessment published on 21 January by the European Centre for Disease Prevention and Control also noted the broad diagnostic criteria and said: “It is expected that many of the suspected cases will be reclassified and discarded.”

The ECLAMC’s conclusion — that the apparent surge could be largely an artefact — is possible in principle, says Helen Dolk, an epidemiologist at the University of Ulster near Belfast, UK, who works on the surveillance of congenital abnormalities. But she stresses that it is impossible to confirm until more data become available — and that she is reserving judgement on the portion of the apparent increase that can be attributed to confounding factors.

## REAL CONCERN

By contrast, Lavinia Schuler-Faccini, a researcher at the Federal University of Rio Grande do Sul, and president of the Brazilian Society of Medical Genetics, says she is certain that there has been a substantial increase in microcephaly cases. She notes that physicians began reporting a rise before the increased attention by health authorities, and the media began reporting a spike last November. “My personal impression is that there is an augmentation of cases of microcephaly in Brazil,” she says, but adds that it is not as huge as the suspected number of cases reported to the health ministry. “All our efforts now are to establish the real level of increase.”

Establishing whether there is a link between microcephaly and Zika is particularly important because people living in the Americas lack immunity to the virus. This, combined with the fact that the *Aedes* mosquitoes that transmit Zika are widespread in the Americas, means that many people will be infected in future, including pregnant women.

Zika virus has been found in amniotic fluid, placental or fetal tissues in several cases of nervous-system malformations, including microcephaly, in Brazil. In a 1 February statement, the WHO said: “A causal relationship between Zika infection during pregnancy and microcephaly is strongly suspected, though not yet scientifically proven.” ■ SEE EDITORIAL P.5

SOURCE: ECDC (ZIKA-TRANSMISSION DATA); M. U. G. KRAEMER ET AL. *ELIFE* 4, E08347; 2015 (MOSQUITO RANGE)



### TOP NEWS



Scientists swab C-section babies with mothers’ microbes  
[go.nature.com/houq7r](http://go.nature.com/houq7r)

### MORE ONLINE

- ArXiv rejections lead to spat over screening process [go.nature.com/w4p2qh](http://go.nature.com/w4p2qh)
- Error found in study of first ancient African genome [go.nature.com/ks4jeo](http://go.nature.com/ks4jeo)
- Babylonian astronomers used geometry to track Jupiter [go.nature.com/w9omzk](http://go.nature.com/w9omzk)

### NATURE PODCAST



Killing old cells to lengthen life, forests as carbon sinks, and brain-tickling comedy  
[nature.com/nature/podcast](http://nature.com/nature/podcast)



**CORRECTION**

The two elements in the key for the map in the News story 'Microcephaly surge in doubt' (*Nature* **530**, 13–14; 2015) were accidentally swapped round. The corrected graphic can be seen at [go.nature.com/ekmksk](http://go.nature.com/ekmksk).

## POLITICS

# Academics across Europe join 'Brexit' debate

*If the United Kingdom leaves the EU, researchers throughout the bloc will feel the effects.*

BY DANIEL CRESSEY

A UK debate over whether to leave the European Union (EU) is drawing in academics from across the continent. Millions in research funding, collaborations and the employment status of thousands of scientists could be affected by the outcome. The UK government has committed to holding a referendum on a Brexit — as the possible British exit has been dubbed — by the end of 2017.

Brexit proponents say that a lone United Kingdom would enjoy greater freedom to set policies that affect research. But most researchers who are willing to pick a side publicly are worried about the prospect. “Europe would suffer and the UK would suffer,” says Lesley Wilson, secretary-general of the European University Association in Brussels.

One big uncertainty is money. UK universities rely on the EU for around 16% of their total research funding. And scientists working in the United Kingdom are disproportionately successful at winning such awards compared to applicants in other member states; under the EU’s last Framework Programme, which ran from 2007 to 2013, they won grants worth €7 billion (US\$7.58 billion), second in value only to Germany. UK institutions also host more researchers with grants from the EU-funded European Research Council (ERC) than those in any other member state.

Being outside the EU does not necessarily preclude involvement in EU programmes. Non-members such as Norway, Switzerland and Israel have gained access to various EU research schemes, mainly by paying for inclusion in them and by adopting some general EU rules, such as freedom of movement.

Angus Dalglish, who is part of the ‘Leave.eu’ campaign, says that the shortfall would be made up if the United Kingdom were to redirect to science even a portion of the sum — most recently €140 billion — that it must contribute annually to the overall budget of the EU as a member. “We would have a far bigger budget for funding our own science,” says Dalglish, a cancer and HIV researcher at St George’s, University of London, who once stood for election as a member of the pro-Brexit UK Independence Party.

But opponents of a Brexit counter that the United Kingdom could find it expensive to renegotiate entry into EU funding programmes



The fusion reactor ITER in France could become less accessible to UK scientists in the event of a Brexit.

because of its historically outsized success at winning competitive grants. They also cite the example of Switzerland, which had bought into EU funding schemes but was booted out of many such projects in 2014 after its citizens voted to restrict immigration. Scientists had to scramble to find ways around the official severing of ties after that, and Swiss participation has been restricted ever since.

UK researchers could still access major EU research-infrastructure projects after a Brexit. But they would lose the priority given to EU members, putting them at the back of the queue for access to facilities ranging from a laser instrument to a major social-science data set and a huge experimental fusion reactor, ITER.

## BRIDGES BURNED

For Kurt Deketelaere, secretary-general of the League of European Research Universities in Leuven, Belgium, the potential loss of mobility and collaboration is worrying for scientists across Europe. Wilson agrees. “Everybody wants to work with the best,” she says. “If you’re going to lose strong partners, that’s not going to benefit anybody.” Dalglish counters that universities already maintain successful collaborations with non-EU members, and says that opting out would have “no negative impact on scientific collaboration whatsoever”.

Still, around 15% of academic staff at UK institutions are non-UK EU nationals, a figure

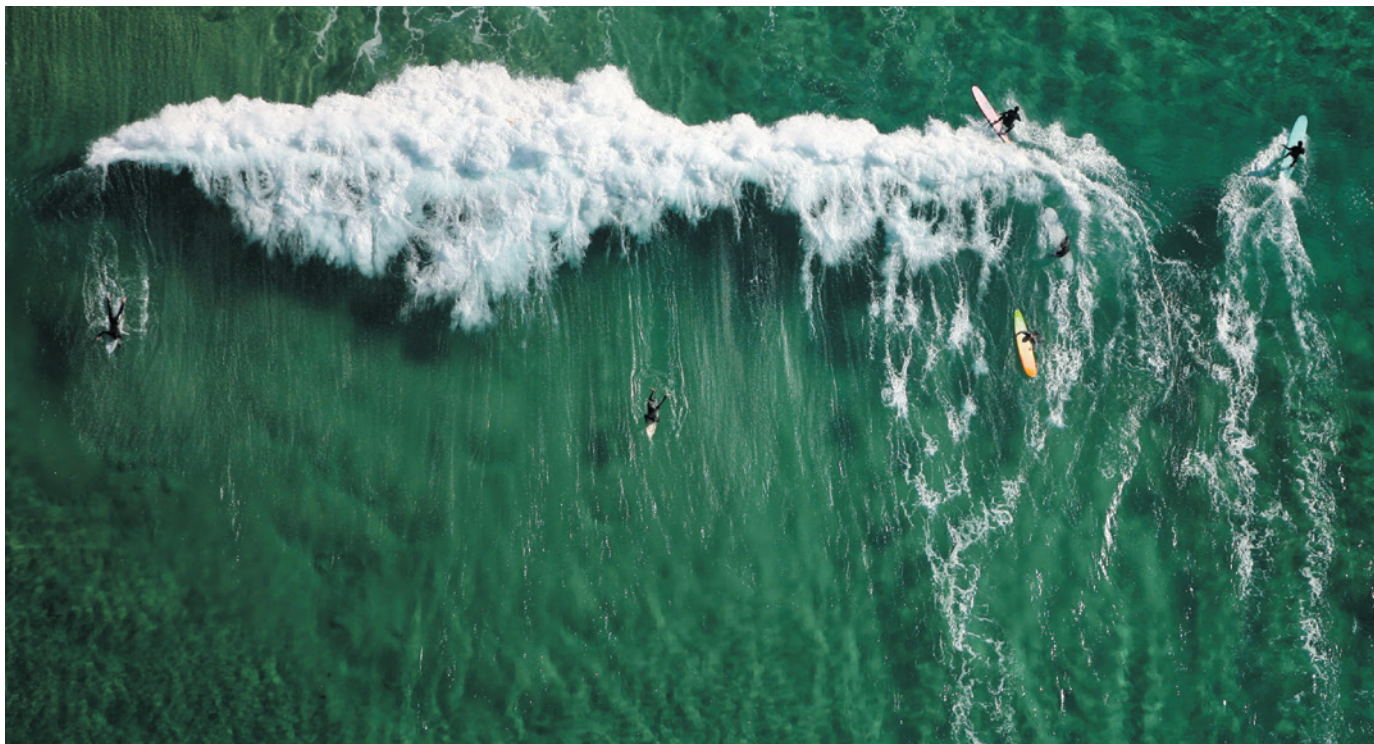
that rises to 20% among elite universities. It is unlikely that these people would have to apply for visas or leave in the event of a Brexit, or that grants already awarded would be clawed back. But in future, obtaining funding and securing jobs across UK–EU borders could become more difficult. “In general, the UK would be less attractive for me,” says Yvonne Peters, a particle physicist originally from Germany who works at the University of Manchester, UK, and is partially funded by a grant from the ERC.

Brexit proponents say that if the United Kingdom leaves, it will escape other EU regulations, such as rules governing clinical trials. These have been widely blamed for hindering UK medical research — triggering proposals to overhaul the regulations. The nation might also be able to offer more tax credits for research spending and adopt a more positive stance on genetically modified crops.

But the EU would also lose a powerful political voice pushing for science, says Vicky Ford, a UK Conservative Party member of the European Parliament who supports EU reform. She says that UK votes have been crucial in parliamentary decisions that have affected the entire bloc, including reductions to red tape in the awarding of funds, and improvements to science-advice mechanisms in EU politics. Without UK votes, Ford says, EU research would have paid a price: “Definitely there would have been less money.” ■ **SEE EDITORIAL P.6**

ITER ORGANIZATION





STEVE CHRISTO/FARFAX MEDIA/GETTY

Surfing and science have joined forces as the World Surf League, the sport's governing body, partners with a climate-research centre at Columbia University.

## CLIMATE RESEARCH

# University calls on private donors for climate funding

*Columbia courts philanthropic benefactors to support research on impacts and adaptation.*

BY JEFF TOLLEFSON

Like most researchers at Columbia University's Lamont–Doherty Earth Observatory (LDEO) in Palisades, New York, Park Williams is expected to win research grants to cover his salary. But times are tough for climate scientists, who face flat levels of government funding in an ever-expanding pool of competitors. Two years into a post as an assistant research professor, the 34-year-old bioclimatologist had yet to receive a single grant.

But on 22 January, the Center for Climate and Life, a new research institute at Columbia that is seeking funds from the business community, awarded Williams US\$180,000 for his work on historical drought and fire cycles. Michael Puma, an environmental modeller at Columbia and NASA's Goddard Institute for Space Studies in New York City, received \$190,000 to explore the impact of climate change on the global food system.

"I was beginning to worry about my future here and wondering if I'd be wise to begin

applying for more-traditional professorships," says Williams. "Now I'll have time to actually do research."

With climate-science funding under

perennial threat in Washington DC, Columbia is engaging corporate philanthropists to boost research into the effects of projected environmental changes and how human systems can adapt. Seeded by Columbia with an initial budget of \$3.1 million over five years, the Center for Climate and Life hopes to build a \$200-million endowment that disburses around \$10 million annually.

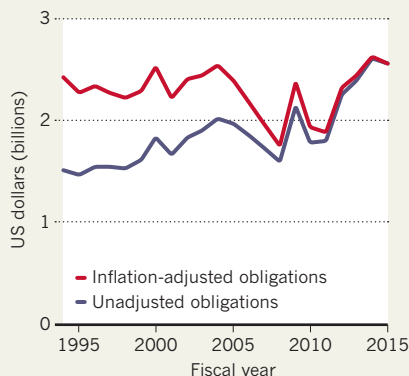
"It's a very new way of funding science," says Peter de Menocal, a palaeoclimatologist at the LDEO and founding director of the new centre.

The centre will supplement salaries and research costs for scientists at Columbia, as well as at NASA's Goddard Institute for Space Studies, which is located on the Columbia campus. De Menocal says that the centre will apply the same peer-review procedures used by the US National Science Foundation to ensure that its grants are directed towards the best research proposals.

With roughly 85% of scientists at the LDEO reliant on government grants for their salaries, de Menocal says, two decades of stagnant

## GOING NOWHERE

When adjusted for inflation, US government spending on environmental-science research has not increased substantially in 20 years.



SOURCE: NATIONAL SCIENCE FOUNDATION

budgets for the environmental sciences have taken a toll (see 'Going nowhere'). And Republicans in the House of Representatives, many of whom deny the reality of global warming, have attempted repeatedly to cut funding for climate-related research. De Menocal says that the situation has many young Earth scientists rethinking their career choice because of what he calls "a silly ideological divide." The new institute should give them needed job security, he says, and allow the community to identify and pursue new research paths without waiting for Washington to come around.

Others have also recently turned to private philanthropy to fund climate research. "It is going to be very hard for the government to undertake a really big increase in federal research," says Margaret Leinen, director of the Scripps Institution of Oceanography in La Jolla, California. In August, Scripps opened the Center for Climate Change Impacts and Adaptation, with a donation of \$5 million from energy executive Richard Hertzberg and his wife Carol Dean Hertzberg. And the Grantham Foundation for the Protection of the Environment, founded by investment manager Jeremy Grantham and his wife Hannelore, has helped to establish similar research institutes at multiple universities, including the Grantham Research Institute on Climate Change at the London School of Economics.

One of the Columbia centre's initial partners is the World Surf League (WSL) in Santa Monica, California, which is the governing body of professional surfing. The organization says that it seeks to promote environmental awareness among more than 120 million surfing fans around the world. As part of that partnership, Columbia plans to develop an online certificate programme focused on ocean science and conservation. The courses will be open to anybody beginning in 2017, and may evolve into a formal master's degree programme.

"We've got this perfect combination of science and soul," says Scott Hargrove, chief marketing officer for the WSL, which plans to announce its funding commitment as early as this month. "Surfing has the power to move culture," he says, with Columbia driving the science and education, and surfers serving as public ambassadors.

De Menocal is also in talks with French aerospace giant Airbus, which would provide the fuselage for a research aircraft that could be readily equipped with instruments to study everything from the atmosphere to rainforests and polar ice sheets. De Menocal says that the centre is currently working on a viability study for the project.

"We want to change the way we do and fund science," says de Menocal, "and fast track the science we need to understand how climate impacts people." ■

## PSYCHOPHARMACOLOGY

# Drugmakers target fog of depression

*Industry and researchers push for reliable ways to assess cognitive deficits associated with mood disorders.*

BY HEIDI LEDFORD

In the past quarter of a century, a wave of drugs has transformed the treatment of depression. But the advances have struggled to come to grips with symptoms that often linger long after people start to feel better: cognitive problems such as memory loss and trouble concentrating.

On 3 February, the US Food and Drug Administration (FDA) will convene a meeting of its scientific advisers to discuss whether such cognitive impairments are components of the disorder that drugs might be able to target — or just a result of depressed mood. The discussion will help the agency to decide whether two companies that sell the antidepressant vortioxetine should be allowed to label it as a treatment for the cognitive effects. A 'yes' could spur drug developers to invest in ways to test cognitive function during their antidepressant trials.

Psychiatrists have long noted that some people with depression also struggle to concentrate and to make decisions. The question has been whether such difficulties are merely an offshoot of altered mood and would thus clear up without specific treatment, says Diego Pizzagalli, a neuroscientist at McLean Hospital, an affiliate of Harvard Medical School in Belmont, Massachusetts.

But some patients who report improved mood after treatment still struggle with cognitive deficits — so psychiatrists sometimes prescribe concentration-enhancing drugs that are approved to treat attention deficit hyperactivity disorder to people with depression.

The scenario is a familiar one for those who treat schizophrenia: antipsychotic drugs may drive away hallucinations, but the cognitive deficits persist. And the deficits make it difficult for people with schizophrenia to keep jobs or to live independently, says Michael Green, a neuropsychologist at the University of California, Los Angeles.

## LONG LEAD TIME

More than a decade ago, companies waged a campaign to encourage drug regulators to recognize cognitive impairment in schizophrenia. But the FDA refused to do so until drugmakers came up with uniform criteria to measure the impairments. As a result, the

schizophrenia community built a consensus around a battery of tests for use in clinical trials. In the case of depression, tests would have to be especially sensitive because the cognitive impairments can be more subtle than those that accompany schizophrenia, says Richard Keefe of the Duke Institute for Brain Sciences in Durham, North Carolina.

Researchers and industry representatives discussed the problem of cognitive impairment in depression at a workshop held by the US Institute of Medicine (now the National Academy of Medicine) almost a year ago, but they did not set a course for establishing uniform assays. And even if tests acceptable to the FDA can be established, that is no guarantee that effective drugs will soon follow. Guidelines governing schizophrenia trials were established in 2005, but no cognitive-function drug has yet been approved in such cases. Furthest along is the company Forum Pharmaceuticals in Waltham, Massachusetts, which is conducting late-stage clinical trials of encenicline — a drug that targets the memory-related nicotinic protein acetylcholine receptor  $\alpha 7$ . Results are expected in the first half of this year.

But interest in cognitive drugs for people with depression is building as more and more antidepressants become available in cheap, generic forms and pharmaceutical companies seek to carve out niches for their newer, more expensive offerings, says psychiatrist Eduard Vieta at Spain's University of Barcelona. "Companies are changing strategies, and trying to find indications that are not the typical ones," he says. "When you can speak about an indication that nobody else has — like cognition in the context of depression — it's a huge advantage."

In making the case for vortioxetine, Takeda Pharmaceutical Company in Osaka, Japan, and H. Lundbeck in Valby, Denmark, cite clinical-trial data showing that the drug led to improvements in several cognitive tests, apart from its effect on mood.

If the FDA does decide to recognize cognitive dysfunction as a treatable aspect of depression, the effects could also reach beyond the pharmaceutical industry, says Green. "It's a matter of respecting an aspect of the illness that we've always thought wasn't getting enough attention," he says. "The more visibility there is on these deficits, the better." ■



## BIOTECHNOLOGY

# Embryo editing gets green light

*UK decision sets precedent for research on editing genomes of human embryos.*

BY EWEN CALLAWAY

A team of scientists in London is preparing to edit the genomes of human embryos for research. The UK Human Fertilisation and Embryology Authority (HFEA) approved the work on 1 February in the world's first such endorsement. "It's an important first," says George Daley, a stem-cell biologist at Boston Children's Hospital in Massachusetts. "This establishes a strong precedent for allowing this type of research."

The successful applicant is developmental biologist Kathy Niakan, at the Francis Crick Institute in London. Her team plans to use the CRISPR-Cas9 technique in healthy human embryos to alter genes that are active just after fertilization. The researchers will stop the experiments after seven days, when the embryos will be destroyed.

The genetic modifications could lead to treatments for infertility, but will not themselves form the basis of a therapy.

Robin Lovell-Badge, a developmental biologist at the Crick institute, says that the HFEA's decision will embolden other researchers. He has heard from other UK scientists who are interested in pursuing embryo-editing research, he says, and expects more applications to follow. Researchers also expect the decision to reverberate beyond the United Kingdom. "I think this will be a good example to countries who are considering their approach to regulating this technology," says bioethicist Sarah Chan at the University of Edinburgh, UK. "We can have a well-regulated system that is able to make that distinction between research and reproduction."

It remains illegal to alter the genomes of embryos used to conceive a child in the United Kingdom. But Tetsuya Ishii, a bioethicist at Hokkaido University in Sapporo, Japan, says that the HFEA decision could also stimulate debate over deploying embryo gene editing in clinical settings.

At a press briefing last month, Niakan said that her team could begin work within "months" of the HFEA approving the application. First,

however, a local research ethics board (similar to an US institutional review board) must approve the work. ■

➔ **NATURE.COM**  
For a longer version of this story, see: [go.nature.com/1sluwv](http://go.nature.com/1sluwv)



Ludwig Maximilian University of Munich is part of a government-backed 'cluster of excellence'.

## GERMANY

# Science hubs win in major revamp

*Research clusters emerge as the big success of Germany's Excellence Initiative — despite its focus on elite institutes.*

BY QUIRIN SCHIERMEIER

For many, Munich's fame rests on the Oktoberfest beer festival. But for astrophysicist Stephan Paul, what makes the Bavarian capital so charming is its universities' rise to stardom in studies on the origin and structure of the Universe.

The region has long been a national hub for physics, but its appeal to theorists and particle physicists has soared in recent years thanks to a well-funded research programme that brings together the city's two large universities — the Technical University of Munich (TUM) and Ludwig Maximilian University (LMU) — and several Max Planck institutes in nearby Garching.

"The research infrastructure here is top-notch and the concentration of expertise is quite unique," says Paul, a physicist at the TUM who coordinates the programme.

The programme is one of 43 'clusters of excellence' launched in 2011 as part of Germany's €4.6-billion (US\$5-billion) Excellence Initiative. The clusters are among the ten-year-old initiative's most tangible

successes, according to a major report released on 29 January by an independent, international panel (see [go.nature.com/qxo768](http://go.nature.com/qxo768)).

The hubs bring together research groups — either within a university or across different institutes in the same region — that previously had little contact, so that they can pool their facilities and build on each other's successes. "We were surprised to find out how much good science there was just around the corner," says Paul.

The report, commissioned by Germany's federal government and its 16 state governments, strongly recommends that they continue the excellence initiative, in particular the highly successful clusters. The report is less conclusive on the initiative's success in achieving its much higher-profile goal: to produce a top-ranked research powerhouse akin to Harvard University or the universities of Oxford or Cambridge.

"The high quality of science produced at the clusters of excellence is particularly impressive," says Dieter Imboden, a Swiss environmental physicist and long-time science manager who chaired the evaluation panel.

SUEDEUTSCHE ZEITUNG PHOTO/ALAMY

SOURCE: IJKE FINAL REPORT (JAN. 2016);  
HTTP://GO.NATURE.COM/QX0768

“But we are only at the beginning of a long road towards the group of global top universities.”

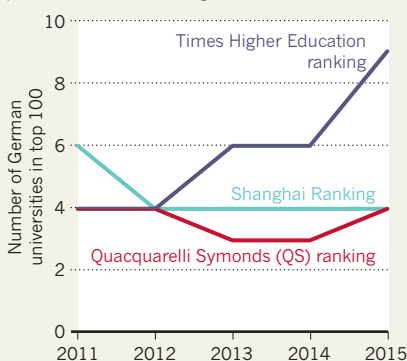
Since it began, the initiative has designated a few universities as ‘elite’ — the latest assignment gave the status to 11 universities, including the TUM and the LMU — and rewarded them with an extra €10 million to €14 million per year. A report published last September by Germany’s main research-funding agency, the DFG, noted that these 11 universities have markedly increased their scientific output. A further analysis done last year by *Nature* (*Nature* 525, 168–169; 2015) found that at elite institutions, the proportion of publications that feature in the top 10% of the world’s most highly cited papers had almost doubled since 2002 (although *Nature* also found that the same was true on average of five good, but not elite, German universities).

But the new elite universities still lag far behind the likes of Oxford and Harvard in terms of world rankings, appeal to top scientists and funding. Critics of the initiative say that it has created a two-tier research system and an excessive administrative burden.

Even scientists at the ‘elite’ universities agree that the jury is still out on the success of the concept (see ‘Germany’s elite?’). “The initiative has created a palpable ‘we can do it’ spirit,” says Stephan Leibfried, a social scientist and research-policy specialist at the University of

## GERMANY’S ELITE?

A batch of ‘elite’ German universities received extra funds in 2011 — but the effect on the country’s presence in world rankings is inconclusive.



Bremen, which has elite status. But reaching the top of international university rankings “will require decades of hard work”.

“No matter how often you might try, a Harvard can’t be pulled, like a rabbit, out of the hat,” agrees Stefan Hornbostel, a science-policy researcher at the Humboldt University of Berlin, another of the latest 11 universities to be labelled elite.

The report recommends a two-year extension of the current programme, which ends in 2017, followed by a new programme starting in 2019.

It suggests that no more than 10 institutions should receive the elite bonus in the future, and that prolonged support of promising research clusters should be at the core of a renewed excellence programme. The report also says that in the follow-up regime, “smaller disciplines” deserve proportionally more funding than in the past and that research clusters could span geographically distant universities.

The recommendations are non-binding but are sure to feed into the decision of federal and state governments on whether to continue the initiative, which is due by June.

For Paul and his team, who have received some €70 million from the programme so far, the prospect of renewed support comes at the right time. Since 2006, some 150 physicists have moved from leading institutes in Europe and overseas to Munich. Researchers there have started to operate a newly built ultracold neutron facility and an underground laboratory largely shielded from background radiation, both financed in part with money from the initiative. Key experiments at these new facilities, says Paul, might shed light on the properties of neutrino particles, the design of the early Universe and the elusive nature of dark matter.

The collaborations enabled by the Excellence Initiative have produced some “good friction”, he says. “Now we need to start using the heat.” ■



**CORRECTION**

The dollar conversion of €4.6 billion in the News story 'Science hubs win in major revamp' (*Nature* **530**, 18–19; 2016) should have been US\$5 billion not \$5 million.



# THE MYSTERY OF THE EXPANDING TROPICS

AS EARTH'S DRY ZONES SHIFT RAPIDLY POLEWARDS, RESEARCHERS ARE SCRAMBLING  
TO FIGURE OUT THE CAUSE — AND CONSEQUENCES.

BY OLIVE HEFFERNAN



**Severe droughts over the past 20 years in Australia have taken a toll on farmers, who struggle to keep livestock and crops healthy.**

California — that's enough to have a huge impact on those communities in terms of how much rain they will get," explains climate modeller Thomas Reichler of the University of Utah in Salt Lake City.

Since Fu and his colleagues announced their discovery<sup>1</sup> in 2006, many scientists have investigated the tropical bloating and tried to decipher its cause. Explanations range from global warming to ozone depletion or natural cycles that will reverse in the future. And there is little agreement on how quickly the border of the tropics is shifting: estimates run from less than half a degree of latitude per decade to several. At the more extreme end, the change in climate would be like moving London to the position of Rome over the course of a century<sup>2-5</sup>. The problem is compounded by lack of consensus on how to define the tropics, which makes it hard for scientists to agree on the extent of the changes. Nevertheless, researchers investigating this phenomenon agree that it is real.

"There's a big need to be concerned about this issue," says climate scientist Chris Lucas at the Australian Bureau of Meteorology in Melbourne. That's because of the possible impacts: some of the world's most fertile fishing grounds could disappear, global grain production could shrink and biodiversity could suffer.

### STRANGE SKIES

At the same time as Fu first discovered odd patterns in the satellite data, Reichler noticed something unusual in the skies. He was researching the tropopause, the boundary between the lowest level of the atmosphere (the troposphere) and the layer above it (the stratosphere). At the Equator, the tropopause is normally several kilometres higher than at the poles, because warm air rises and pushes the boundary upwards. While analysing temperature data collected from weather balloons, Reichler had found that this equatorial bulge in the tropopause was expanding towards the poles, a sign that the tropics were growing. Fu heard about Reichler's data, and they decided to publish their discoveries together<sup>1</sup>.

Ten years after they sounded the alarm, scientists are still struggling to work out what is happening. Last July, 50 researchers gathered in Santa Fe, New Mexico, to discuss everything that is known about tropical expansion — how to measure it, what is causing it and where the future border of the tropics might be. "We're at a stage where we recognize the problem is more complex than we originally thought," explains the organizer of the conference, Dian Seidel, an atmospheric scientist with the US National Oceanic and Atmospheric Administration (NOAA) in Silver Spring, Maryland.

Some of the changes in the tropics could be a result of global warming. Reichler investigated

that possibility in a study<sup>6</sup> led by Jian Lu, an Earth systems scientist now at the Pacific Northwest National Laboratory in Richland, Washington. Working with Gabriel Vecchi, a climate scientist with NOAA in Princeton, New Jersey, the researchers looked at climate forecasts to see how warming might affect an atmospheric circulation pattern called the Hadley cell, which transports heat from the warmer parts of Earth towards the cooler regions (see 'Bulging waistline'). As part of the Hadley cell, warm, moist air soars skywards above the Equator and cool, dry air tumbles towards Earth at about 30° latitude in the Northern and Southern Hemispheres. That downward limb of the Hadley cell helps to create some

## THE CHANGE IN CLIMATE WOULD BE LIKE MOVING LONDON TO THE POSITION OF ROME OVER THE COURSE OF A CENTURY.

of the driest deserts on the planet, such as the Kalahari in southern Africa and the Sahara in northern Africa, and it is one of the most common measures of the boundary between the tropics and the drier subtropics.

In their study, Lu and his colleagues found that climate models generally forecast that the outer edge of the Hadley cell will shift because of global warming. But the models predict a much slower rate of tropical expansion than has been seen so far — which has led researchers to suspect that something else is going on.

A common view, and one held by Lucas, is that natural climatic variability is playing some part. That variability could take the form of large-scale climatic cycles such as the Pacific Decadal Oscillation, in which temperatures in the Pacific Ocean swing between hot and cold across time-scales of 15–20 years or more. "Or it could be in the form of much more random, chaotic noise," says Lucas, who thinks that large cycles and noise together account for 50% or more of the expansion. Atmospheric scientist Darryn Waugh at Johns Hopkins University in Baltimore, Maryland, agrees. "It's a chaotic system, so some of the variability is just noise in the system." If that is the case, tropical expansion could slow down or even reverse in some regions when those natural variations swing back.

Another answer might involve different forces in the Northern and Southern hemispheres. South of the Equator, tropical expansion has been strongest in the summer, and that leads some researchers to suspect that it is related to the pattern of ozone loss in the southern stratosphere. Pollutants chew up ozone molecules above Antarctica in the spring, which triggers circulation changes throughout other parts of the Southern Hemisphere during summer. The correlation with tropical expansion suggests that the two

One spring day in 2004, Qiang Fu was poring over atmospheric data collected from satellites when he noticed an unusual and seemingly inexplicable pattern. In two belts on either side of the equator, the lower atmosphere was warming more than anywhere else on Earth. Fu, an atmospheric scientist at the University of Washington in Seattle, was puzzled.

It wasn't until a year later that he realized what he had discovered: evidence of a rapid expansion of the tropics, the region that encircles Earth's waist like a green belt. The heart of the tropics is lush, but the northern and southern edges are dry. And these parched borders are growing — expanding into the subtropics and pushing them towards the poles.

Cities that currently sit just outside the tropics could soon be smack in the middle of the dry tropical edge. That's bad news for places like San Diego, California. "A shift of just one degree of latitude in southern

phenomena could be connected. What's more, climate models that factor in ozone loss are able to account for much more of the tropical expansion between 1980 and 2000, when the Antarctic ozone hole was growing bigger nearly every year, says Waugh.

In the Northern Hemisphere, a different explanation is called for because, in general, the Arctic does not suffer the same sort of ozone loss as the Antarctic. Research led by climate scientist Bob Allen at the University of California, Riverside, suggests that the culprits in the north might be black soot and tropospheric ozone — which are both generated by burning fossil fuels. Allen and his team ran simulations with a climate model that featured detailed atmospheric physics, and their analysis showed that black soot and tropospheric ozone have heated the atmosphere in the Northern Hemisphere and driven tropical expansion more than carbon dioxide and other greenhouse gases, particularly in summer<sup>8</sup>.

Not everyone is comfortable with the idea that entirely separate factors could drive tropical expansion to such a large extent on either side of the Equator. Fu, for one, thinks it's unlikely given the similar patterns in the north and south. "If ozone depletion was dominating the expansion in the Southern Hemisphere in the past 30 years, would you see such symmetry? I'm not convinced," says Fu.

The proliferation of hypotheses shows how much researchers are struggling to explain what's happening. "I think we're piecing this together slowly," says Lucas. "We don't have a full explanation yet and I don't think there's going to be one single explanation. It's going to be a little bit of this and a little bit of that."

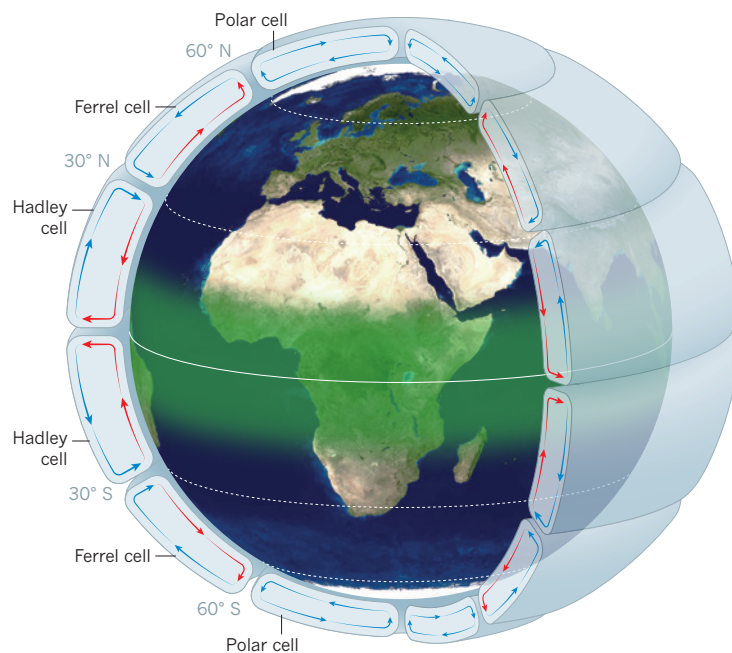
## EDGE EFFECT

Right around the time that scientists were first warning about tropical expansion, Lucas was experiencing what might have been its effects first-hand. During 2006 and 2007, Australia was deep in the middle of one of the worst droughts to have hit the continent since Europeans settled there. Lucas recalls driving from Melbourne to nearby Lake Eildon and seeing the once-brimming lake empty. Meanwhile, Melbourne's reservoirs were running low, and north of the city, forest fires raged in the mountains. The worst-affected regions of Australia — cities such as Perth, Adelaide and Melbourne — were south of 30° latitude, which suggests that the drying could be caused by a shift in the position of the Hadley cell and the rain-bearing jet stream. According to research published<sup>9</sup> in 2010, southeastern Australia has been invaded by a drier climate from the north in recent decades, which has greatly reduced rainfall. "We can't say that this is exclusively due to tropical expansion, but it's certainly consistent with tropical expansion," explains Lucas. "And our concern is that southeastern Australia is going to keep getting drier."

Elsewhere, there is evidence that tropical expansion is affecting the ocean. Where the

## BULGING WAISTLINE

Expansion of the tropics can be seen in the Hadley cell, the circulation pattern that carries warm air upwards above the Equator and then down at about 30° N and 30° S. The descending limb of each Hadley cell is shifting towards the pole in both hemispheres, potentially altering climatic conditions in some regions.



Hadley cell descends, bringing cool air downward, it energizes the ocean and whips up currents to high speeds. This energy powers the upwelling of cold, nutrient-rich waters towards the surface, which feeds some of the world's most productive fisheries. But there are hints that some of these regions are suffering because of shifts in the Hadley cell.

Edward Vizzy and Kerry Cook<sup>10</sup>, both at the University of Texas at Austin, have found some unhealthy signs in the region of the Benguela Current, an area of coastal upwelling along the coast of west Africa and south of 30° latitude. According to Cook, the currents of that entire region have shifted over the past 30 years. One effect is that the upwelling has weakened, with worrying implications for the region's fisheries and biodiversity. Cook says that the same could be true of open-ocean upwelling systems, which are more susceptible to changes in the position of the Hadley cell.

These upwelling zones could move south over time, or get weaker or stronger, depending on what happens to the Hadley cell, says Cook. In any case, it means that fishing communities that rely on these resources will not be able to count on traditional patterns.

On land, biodiversity is also potentially at risk. This is especially true for the climate zones just below the subtropics in South Africa and Australia, on the southern rim of both continents. In southwestern Australia, renowned as one of the world's biodiversity hotspots, flowers bloom during September, when tourists come to marvel at some of the region's 4,000 endemic plant species. But since the late 1970s, rainfall there has dropped by one-quarter. The same is true at South Africa's Cape Floristic Province, another frontier known for its floral beauty. "This is the most concrete evidence we have of tropical expansion," says Steve Turton, an environmental geographer at James Cook University in Cairns, Australia.

Turton worries that the rate of change will be too rapid for these ecosystems to adapt. "We're

talking about rapid expansion that's within half or a third of a human lifetime," he says. In the worst-case scenario, the subtropics will overtake these ecologically rich outposts and the hotter, drier conditions will take a major toll.

For the scientists working in this field, communicating the threat of tropical expansion will be tricky, given the level of uncertainty. "It's frustrating to see how much work we have left," says Thomas Birner, an atmospheric scientist at Colorado State University in Fort Collins and one of the conveners of the Santa Fe meeting. One outcome of that conference was an agreement that scientists should compare the various metrics for measuring tropical expansion in the hope of agreeing on the best way forward.

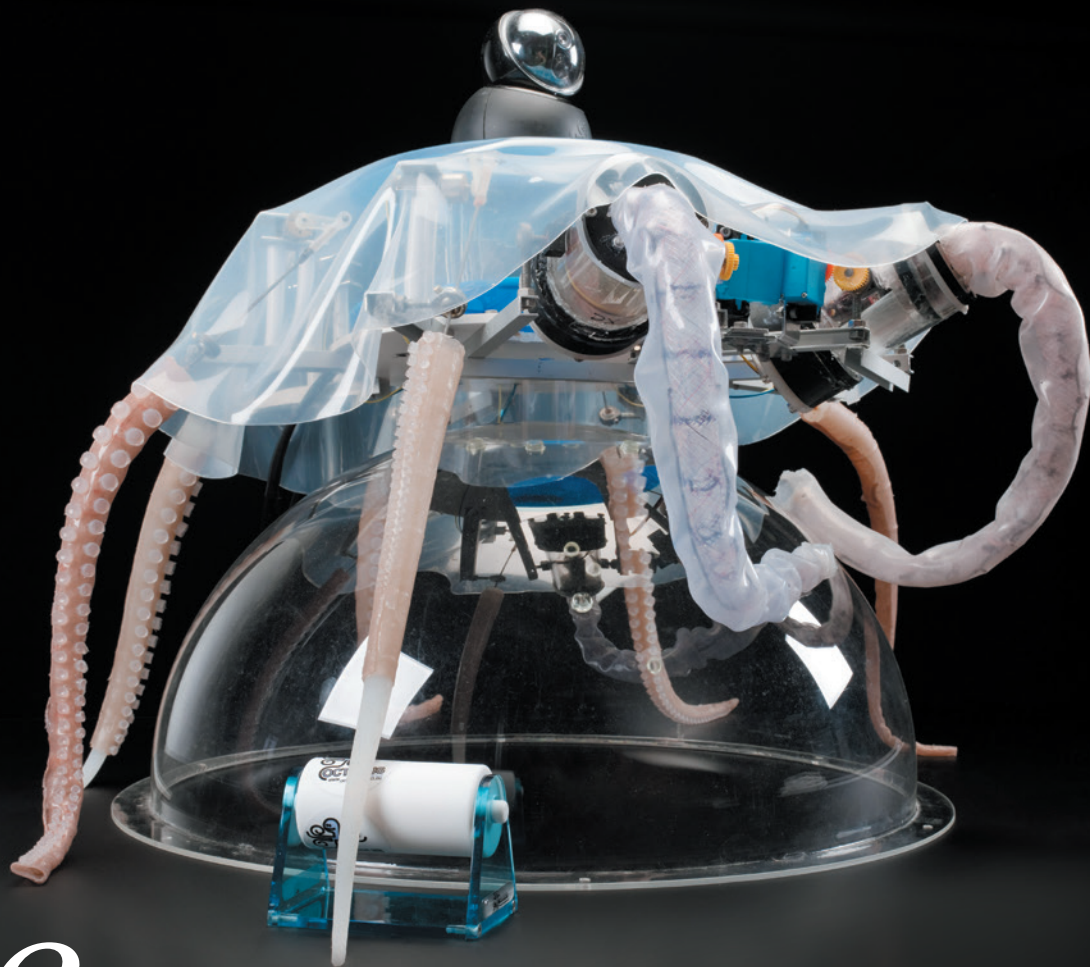
More time will also help. If tropical expansion continues at a fairly constant rate, says Waugh, there will be less of a chance that natural variability is the main culprit, and the finger will point more strongly to other causes.

But that long wait for an answer will be no comfort for the residents of cities such as Santiago, San Diego and Melbourne, and for the billions of others who live near the boundary between the tropics and subtropics. "We need to understand this issue," says Lucas, "to have a sustainable civilization there." ■

**Olive Heffernan** is a writer in Dublin.

1. Fu, Q., Johanson, C. M., Wallace, J. M. & Reichler, T. *Science* **312**, 1179 (2006).
2. Quan, X.-W., Hoerling, M. P., Perlwitz, J., Diaz, H. F. & Xu, T. *J. Climate* **27**, 1999–2013 (2014).
3. Hudson, R. D., Andrade, M. F., Follette, M. B. & Frolov, A. D. *Atmos. Chem. Phys.* **6**, 5183–5191 (2006).
4. Seidel, D. J. & Randel, W. J. *J. Geophys. Res.* **112**, D20113 (2007).
5. Lucas, C., Timbal, B. & Nguyen, H. *WIREs Clim. Change* **5**, 89–112 (2014).
6. Lu, J., Vecchi, G. A. & Reichler, T. *Geophys. Res. Lett.* **34**, L06805 (2007).
7. Waugh, D. W., Garfinkel, C. I. & Polvani, L. M. *J. Climate* **28**, 6581–6586 (2015).
8. Allen, R. J., Sherwood, S. C., Norris, J. R. & Zender, C. S. *Nature* **485**, 350–354 (2012).
9. Nicholls, N. *Clim. Dynamics* **4**, 835–845 (2010).
10. Vizzy, E. K. & Cook, K. H. *Clim. Dynam.* <http://dx.doi.org/10.1007/s00382-015-2691-1> (2015).





# The soft touch

BY HELEN SHEN

**Rigid robots step aside — a new generation of squishy, stretchy machines is wiggling our way.**

**I**n 2007, Cecilia Laschi asked her father to catch a live octopus for her seaside lab in Livorno, Italy. He thought she was crazy: as a recreational fisherman, he considered the octopus so easy to catch that it must be a very stupid animal. And what did a robotics researcher who worked with metal and microprocessors want with a squishy cephalopod anyway?

Nevertheless, the elder Laschi caught an octopus off the Tuscan coast and gave it to his daughter, who works for the Sant'Anna School of Advanced Studies in Pisa, Italy. She and her students placed the creature in a saltwater tank where they could study how it grasped titbits of anchovy and crab. The team then set about building robots that could mimic those motions.

Prototype by prototype, they created an artificial tentacle with internal springs and wires that mirrored an octopus's muscles, until the device could undulate, elongate, shrink, stiffen and curl in a lifelike manner<sup>1</sup>. "It's a completely different way of building robots," says Laschi.

This approach has become a major research front for robotics in the past ten years. Scientists and engineers in the field have long worked on hard-bodied robots, often inspired by humans and other animals with hard skeletons. These machines have the virtue of moving in mathematically predictable ways, with rigid limbs that can bend and straighten only around fixed joints. But they also require meticulous programming and extensive feedback to avoid smacking into things; even then, their motions often become erratic or even dangerous when dealing with humans, new objects, bumpy terrain or other unpredictable situations.

Robots inspired by flexible creatures such as octopuses, caterpillars or fish offer a solution. Instead of requiring intensive (and often imperfect) computations, soft robots built of mostly pliable or elastic materials can just mould themselves to their surroundings. Although

SCIENCE ILLUSTRATION BY JAMES GIBSON

**A robot octopus can move like the real thing.**

some of these machines use wires or springs to mimic muscles and tendons, as a group, soft robots have ditched the skeletons that defined previous robot generations. With nothing resembling bones or joints, these machines can stretch, twist, scrunch and squish in completely new ways. They can transform in shape or size, wrap around objects and even touch people more safely than ever before.

Building these machines involves developing new technologies to animate floppy materials with purposeful movement, and methods for monitoring and predicting their actions. But if this succeeds, such robots might be used as rescue workers that can squeeze into tight spaces or slink across shifting debris; as home health aides that can interact closely with humans; and as industrial machines that can grasp new objects without previous programming.

Researchers have already produced a wide variety of such machines, including crawling robotic caterpillars<sup>2</sup>, swimming fish-bots<sup>3</sup> and undulating artificial jellyfish<sup>4</sup>. On 29–30 April, ten teams will compete in Livorno in an international soft-robotics challenge — the first of its kind. Laschi, who serves as scientific coordinator for the European Commission-backed sponsoring research consortium, RoboSoft, hopes that the event will drive innovation in the field.

“If you look in biology, and you ask what Darwinian evolution has coughed up, there are all kinds of incredible solutions to movement, sensing, gripping, feeding, hunting, swimming, walking and gliding that have not been open to hard robots,” says chemist George Whitesides, a soft-robotics researcher at Harvard University in Cambridge, Massachusetts. “The idea of building fundamentally new classes of machines is just very interesting.”

## SMOOTH MOVES

The millions of industrial robots around the world today are all derived from the same basic blueprint. The metal-bound machines use their hefty, rigid limbs to shoulder the grunt work in car-assembly lines and industrial plants with speed, force and mindless repetition that humans simply can't match. But standard robots require specialized programming, tightly controlled conditions and continuous feedback of their own movements to know precisely when and how to move each of their many joints. They can fail spectacularly at tasks that fall outside their programming parameters, and they can malfunction entirely in unpredictable environments. Most must stay behind fences that protect their human co-workers from inadvertent harm.

“Think about how hard it is to tie shoelaces,” says Daniela Rus, director of the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology in Cambridge. “That’s the kind of capability we’d like to have in robotics.”

Over the past decade, that desire has triggered an increased interest in lighter, cheaper machines that can handle fiddly or unpredictable situations and collaborate directly with humans. Some roboticists, including Laschi, think that soft materials and bioinspired designs can provide an answer.

That idea was a tough sell at first, Laschi says. “In the beginning, very traditional robotics conferences didn’t want to accept my papers,” she says. “But now there are entire sessions devoted to this topic.” Helping to fuel the surge in interest are recent advances in polymer science, especially the development of techniques for casting, moulding or 3D printing polymers into custom shapes. This has enabled roboticists to experiment more freely and quickly with making soft forms.

As a result, more than 30 institutions have now joined the RoboSoft collaboration, which kicked off in 2013. The following year saw the launch of a dedicated journal, *Soft Robotics*, and of an open-access resource called the Soft Robotics Toolkit: a website developed by researchers at Trinity College Dublin and at Harvard that allows researchers and amateurs to share tips and find downloadable designs and other information (see [go.nature.com/8gsq4h](http://go.nature.com/8gsq4h)).

Still, says Rebecca Kramer, a mechanical

engineer at Purdue University in West Lafayette, Indiana, “I don’t think the community has coalesced on what a soft robot should look like, and we’re still picking out the core technology.”

Perhaps the most fundamental challenge is getting the robots’ soft structures to curl, scrunch and stretch. Laschi’s robotic tentacle houses a network of thin metal cables and springs made of shape-memory alloys — easily bendable metals that return to their original shapes when heated. Laid lengthwise along the ‘arm’, some of these components simulate an octopus’s longitudinal muscles, which shorten or bend

the tentacle when they contract. Others radiate out from the tentacle’s core, simulating transverse muscles that shrink the arm’s diameter. Researchers can make the tentacle wave — or even curl around a human hand — by pulling certain combinations of cables with external motors, or by heating springs with electrical currents.

**“It’s a completely different way of building robots.”**

A similar system helps to drive the soft-robotic caterpillars that neurobiologist Barry Trimmer has modelled on his favourite experimental organism, the tobacco hornworm (*Manduca sexta*).

At his lab at Tufts University in Medford, Massachusetts, 20 hornworms are born each day, and Trimmer 3D prints a handful of robotic ones as well. The mechanical creatures wriggle along the lab bench much like the real ones, and they can even copy the caterpillar’s signature escape move: with a pull here and a tug there on the robot’s internal ‘muscles’, the machine snaps into a circle that wheels away<sup>5</sup>. Trimmer, who is editor-in-chief of *Soft Robotics*, hopes that this wide range of movements could one day turn the robot into an aide for emergency responders that can rapidly cross fields of debris and burrow through rubble to locate survivors of disasters.

Whitesides, meanwhile, is pioneering robots that are powered by air — among them a family of polymer-based devices inspired by the starfish. Each limb consists of an internal network of pockets and channels, sandwiched between two materials of differing elasticity. As researchers pump air into different parts of the robot, the arms (or legs or fingers) inflate asymmetrically and curl. Whitesides’ team has even built one device that can play ‘Mary Had a Little Lamb’ on the piano<sup>6</sup>. One of the team’s four-legged creations has mastered a robot obstacle course: ambling towards an elevated partition with a clearance of about 2 centimetres, the machine drops down and shimmies underneath, demonstrating the potential of soft robots to tackle complex terrains<sup>7</sup>.

## GRABBING MARKET SHARE

Although most soft robots remain in the lab, some of Whitesides’ creations are venturing out to feed industrial demand for adept robotic hands. Conventional grippers require detailed information about factors such as an object’s location, shape, weight and slipperiness to move each of its joints correctly. One system may be specialized for handling shampoo bottles, whereas another picks up only children’s toys, and yet another is needed for grabbing T-shirts. But as manufacturers update their product lines, and as e-commerce warehouses handle a growing variety of objects, these companies need to swap in customized grippers and updated control algorithms for each different use — often at great cost and delay.

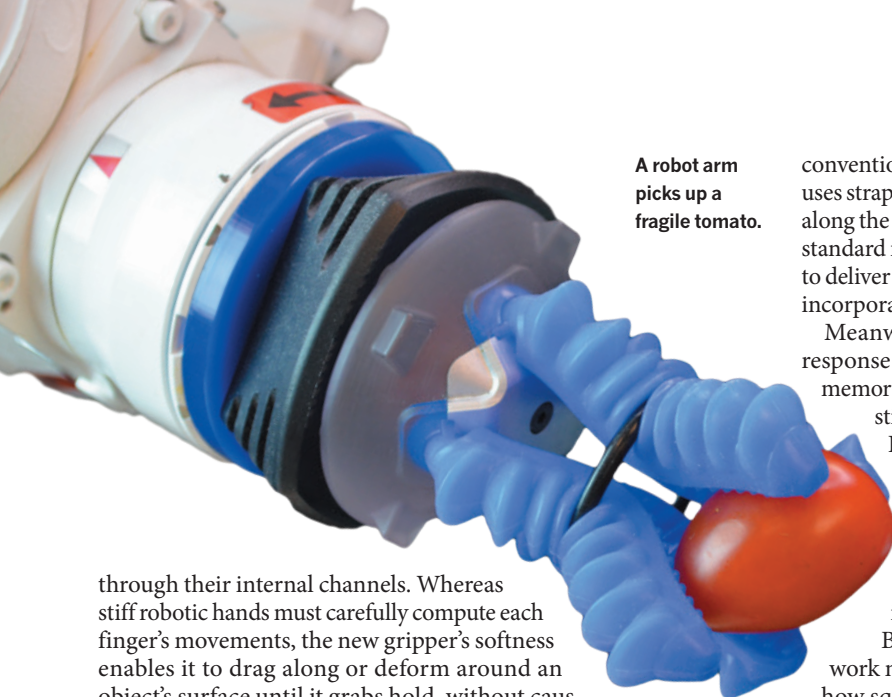
By contrast, grippers that are made mainly of soft, stretchy materials can envelop and conform to objects of different shapes and sizes. Soft Robotics, a start-up company in Cambridge, Massachusetts, that spun out of Whitesides’ research in 2013, has raised some US\$4.5 million to develop a line of rubbery robotic claws. “We use no force sensors, no feedback systems and we don’t do a lot of planning,” says the company’s chief executive, Carl Vause. “We just go and grab an object”, squeezing until the grip is secure.

Made entirely of elastic polymers, the claws curl when air pumps

**➔ NATURE.COM**

For a video of soft robots in action, see [go.nature.com/zx5bux](http://go.nature.com/zx5bux)





A robot arm picks up a fragile tomato.

through their internal channels. Whereas stiff robotic hands must carefully compute each finger's movements, the new gripper's softness enables it to drag along or deform around an object's surface until it grabs hold, without causing damage. It can even pick up mushrooms and ripe strawberries, as well as plump tomatoes off a vine — tasks that have historically required the delicate touch of human workers. Soft Robotics released its first gripper for sale in June 2015, and it is running pilot programmes with six client companies involved in packaging and food-handling.

Empire Robotics in neighbouring Boston has taken a radically different approach, by marketing a robotic 'hand' that resembles a squishy stress ball. Sandlike particles inside the ball flow freely at first, allowing it to deform as it presses firmly into an object. Then, a valve sucks air out of the ball so that the grains inside are forced tightly against each other, causing the ball to harden its grip. Based on research<sup>8</sup> by Heinrich Jaeger at the University of Chicago in Illinois, and Hod Lipson at Cornell University in Ithaca, New York, the 'Versaball' can pick up objects in about one-tenth of a second and lift up to about 9 kilograms.

### SENSE OF PLACE

As robotic octopuses, caterpillars, starfish and other malleable machines come to life, some scientists have begun to focus on better ways to control the devices' actions. "We're talking about floppy, elastic materials," says Kramer. "When something moves on one side, you're not quite sure where the rest of the machine is going to end up." That is why many applications will probably require extra sensors to monitor movement. Yet conventional position and force sensors — rigid or semi-rigid electronic components — don't always work well with soft robots that undergo extreme shape changes.

Engineers such as Yong-Lae Park are tackling this problem by developing stretchable electronic sensors. At Carnegie Mellon University in Pittsburgh, Pennsylvania, Park works on gummy patches that contain liquid-metal circuits sandwiched between sheets of silicone rubber. Poured in a variety of patterns, including spirals and stripes, these liquid circuits can be customized to sense when the device is squished or stretched, and in what direction<sup>9</sup>.

"Stretchable sensors can be as sensitive as skin, depending on how you design them. You can tune them to respond to a slight brush of a finger or to a 30-pound weight," says mechanical engineer Robert Shepherd at Cornell, who has developed methods for 3D printing stretch-sensitive 'skins' directly onto soft robots<sup>10</sup>. Alternating layers of conductive and insulating material produce an electrical signal when prodded or pulled.

Stretchy sensors could have an important role in the growing field of wearable robotics. Funded by the US military, Conor Walsh at Harvard University has spent years developing and honing a soft 'exosuit' for soldiers — a comfier analogue to earlier 'Iron Man'-type exoskeletons, meant to help fighters to carry heavy loads over long distances. Users can still feel the device aiding their movement, but walking in the suit feels "pretty natural", says Walsh — a big improvement from

conventional exoskeletons. Instead of bulky, rigid casings, Walsh's suit uses straps made from nylon, polyester and spandex placed strategically along the legs. And a smattering of position and acceleration sensors — standard rigid devices for now — helps to monitor the wearer's gait and to deliver assistance at the optimal times. The next step, says Walsh, is to incorporate stretchy sensors for a softer, more comfortable experience.

Meanwhile, Kramer has created a robotic fabric that moves in response to electrical current<sup>11</sup>. The muslin sheet, which has shape-memory-alloy coils sewn in, can scrunch by up to 60% in length when stimulated. Smart 'threads' keep tabs on the fabric's movements;

Kramer weaves in stretch-sensitive silicone filaments filled with liquid metal. The concept could be used one day for sleeves or cuffs to help injured or elderly people to move.

Kramer also hopes that the material might be used to assemble robots in space. Astronauts could simply drape an active skin around a piece of foam, for example, to turn it into a working robot.

But before soft robots can fly to space, much foundational work must be done on the ground. Relatively little is known about how squishy materials deform in response to external forces, and how movements propagate through soft masses. In addition, most soft robots remain attached or tethered to hard energy sources, such as batteries or compressed-air tanks. Some researchers are already eyeing the potential of biochemical or renewable sources of energy for soft robots.

The RoboSoft challenge in April could help to spur development. There, the entries will be put through their paces: challenges include racing across a sand pit, opening a door by its handle, grabbing a num-

ber of mystery objects and avoiding fragile obstacles under water. The goal, says Laschi, is to demonstrate that soft robots can accomplish some of the same tasks that stiff robots do, as well as others that they cannot.

"I don't think soft robotics is going to replace traditional robotics, but it will be combination of the two in the future," says Laschi. Many researchers think that rigid robots might retain their superiority in jobs requiring great strength, speed

or precision. But for a growing number of applications involving close interactions with people, or other unpredictable situations, soft robots could find a niche.

At Kings College London, for example, Laschi's collaborators are developing a surgical endoscope based on her tentacle technology. And her team in Italy is developing a full-bodied robot octopus that swims by fluid propulsion, and could one day be used for underwater research and exploration. The prototype already pulses silently through a tank in her lab, as the real octopuses swim in the salty waters just outside.

"When I started with the octopus, people asked me what it was for," says Laschi. "I said, 'I don't know, but I'm sure if it succeeds there could be many, many applications'." ■

*Helen Shen is a freelance writer in Sunnyvale, California.*

1. Laschi, C. *et al. Adv. Robotics* **26**, 709–727 (2012).
2. Umedachi, T., Vikas, V. & Trimmer, B. A. in *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems* 4590–4595 (IROS, 2013).
3. Katschmann, R. K., Marchese, A. D. & Rus, D. *Intl Symp. on Experimental Robotics* (ISER, 2014).
4. Villanueva, A., Smith, C. & Priya, S. *Bioinsp. Biomim.* **6**, 036004 (2011).
5. Lin, H.-T., Leisk, G. G. & Trimmer, B. A. *Bioinsp. Biomim.* **6**, 026007 (2011).
6. Mosadegh, B. *et al. Adv. Func. Mater.* **24**, 2163–2170 (2014).
7. Shepherd, R. F. *et al. Proc. Natl Acad. Sci. USA* **108**, 20400–20403 (2011).
8. Brown, E. *et al. Proc. Natl Acad. Sci. USA* **107**, 18809–18814 (2010).
9. Park, Y.-L., Chen, B.-R. & Wood, R. J. *IEEE Sensors J.* **12**, 2711–2718 (2012).
10. Robinson, S. S. *et al. Extreme Mech. Lett.* **5**, 47–53 (2015).
11. Yuen, M., Cherian, A., Case, J. C., Seipel, J. & Kramer, R. K. *IEEE/RSJ Intl Conf. on Intelligent Robots and Systems* 580–586 (IROS, 2014).

# COMMENT

**EXHIBITION** Adolf Fleischmann, pathology sculptor and abstract artist **p.30**

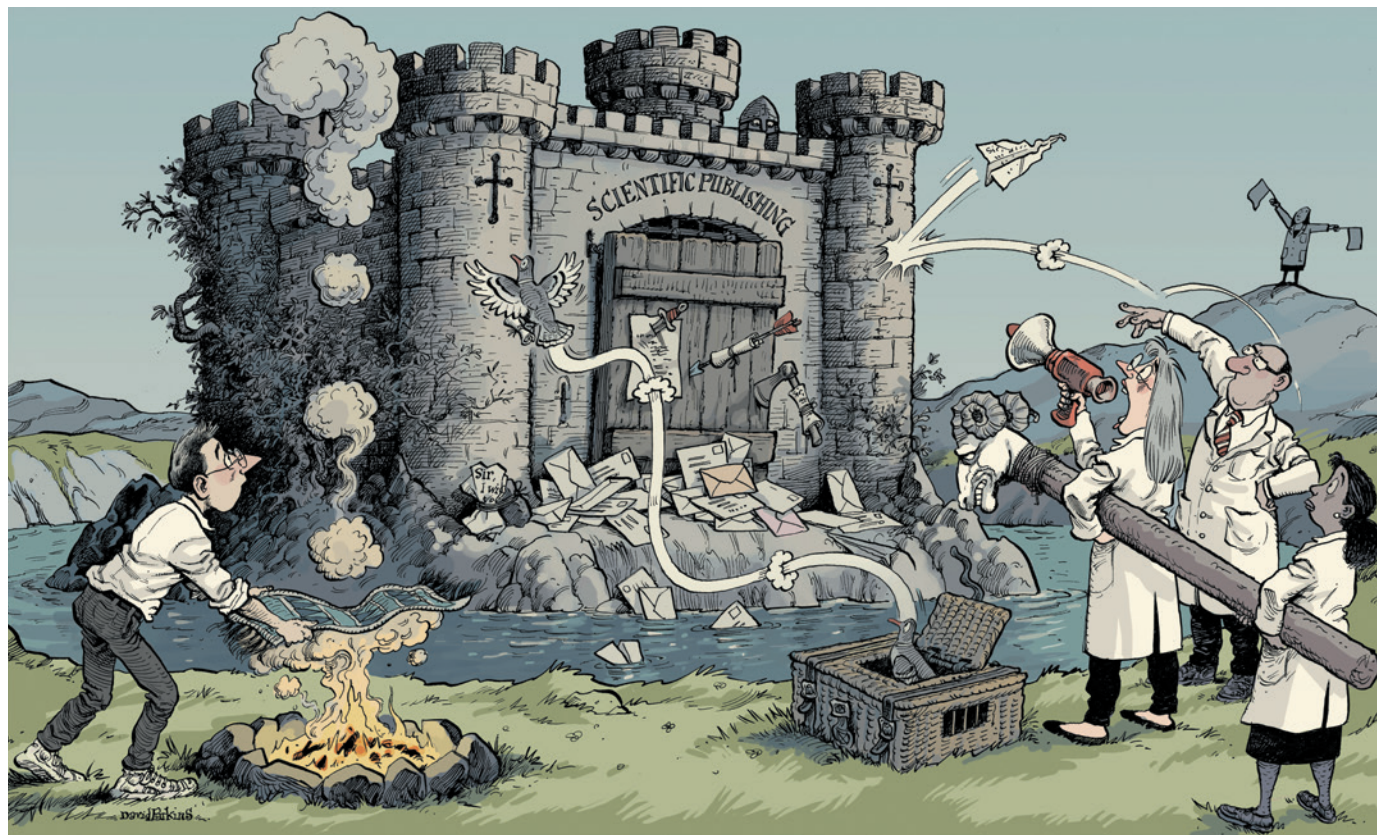


**INTERDISCIPLINARITY** Can architecture catalyse creativity at the Crick? **p.32**

**CONSERVATION** Legal loophole allows mango farmers to cull fruit bats in Mauritius **p.33**

**GEOLOGY** Deep-drilling pioneers: what were they drinking? **p.33**

ILLUSTRATION BY DAVID PARKINS



## A tragedy of errors

Mistakes in peer-reviewed papers are easy to find but hard to fix, report **David B. Allison** and colleagues.

**J**ust how error-prone and self-correcting is science? We have spent the past 18 months getting a sense of that.

We are a group of researchers working on obesity, nutrition and energetics. In the summer of 2014, one of us (D.B.A.) read a research paper in a well-regarded journal estimating how a change in fast-food consumption would affect children's weight, and he noted that the analysis applied a mathematical model that overestimated effects by more than tenfold. We and others submitted a letter<sup>1</sup> to the editor explaining the problem. Months later, we

were gratified to learn that the authors had elected to retract their paper. In the face of popular articles proclaiming that science is stumbling, this episode was an affirmation that science is self-correcting.

Sadly, in our experience, the case is not representative. In the course of assembling weekly lists of articles in our field, we began noticing more peer-reviewed articles containing what we call substantial or invalidating errors. These involve factual

mistakes or veer substantially from clearly accepted procedures in ways that, if corrected, might alter a paper's conclusions.

After attempting to address more than 25 of these errors with letters to authors or journals, and identifying at least a dozen more, we had to stop — the work took too much of our time. Our efforts revealed invalidating practices that occur repeatedly (see 'Three common errors') and showed how journals and authors react when faced with mistakes that need correction.

We learned that post-publication ►

**NATURE.COM**  
For Nature's special collection on reproducibility, see: [go.nature.com/huhbyr](http://go.nature.com/huhbyr)



► peer review is not consistent, smooth or rapid. Many journal editors and staff members seemed unprepared or ill-equipped to investigate, take action or even respond. Too often, the process spiralled through layers of ineffective e-mails among authors, editors and unidentified journal representatives, often without any public statement added to the original article. Some journals that acknowledged mistakes required a substantial fee to publish our letters: we were asked to spend our research dollars on correcting other people's errors.

As academics who publish, review

papers or serve as editors, we appreciate that these issues are complicated. And we feel that journal editors are dedicated and sincere in their efforts. Nevertheless, the scientific community must improve.

Science relies essentially but complacently on self-correction, yet scientific publishing raises severe disincentives against such correction. One publisher states that it will charge the author who initiates withdrawal of a published paper US\$10,000.

Here we summarize our experience, the main barriers we encountered, and

our thoughts on how to make published science more rigorous. (Details of other resolved issues are available on request.)

## SIX PROBLEMS

**Editors are often unable or reluctant to take speedy and appropriate action.** For one paper, we obtained raw data deposited online, received institutional approval to reanalyse the data, and submitted a letter to the editor (through the manuscript-submission system) describing a need for correction within two weeks. After nine months, we asked the journal why, at minimum, an expression of concern had not been posted. An editor admitted that they had not anticipated the process taking as long as it had. The journal communicated its decision to accept our letter and retract the article 11 months after our submission. The letter and retraction have yet to be published.

**Where to send expressions of concern is unclear.** Journals rarely state whom to contact about potentially invalidating errors. We had to guess whether to send letters to a staff member or editor, formally submit the letter as a manuscript, or contact the authors of a paper directly. On a few occasions, we opted to contact authors when an apparent invalidating error may have merely been an ambiguous description. In unequivocal cases, we usually contacted the journal. Often, journals provided no way to contact editors directly, and editorial staff corresponded without identifying themselves; we were unsure whether editors were involved.

**Journals that acknowledged invalidating errors were reluctant to issue retractions.** In one case, we and others found that a paper had mistakenly argued that a statistical adjustment introduced bias, and we submitted a letter to the editor through the journal's submission system<sup>2</sup>. An external statistical review subsequently commissioned by the journal confirmed the error. The authors were asked to retract the article, but they refused. The journal ultimately posted the authors' response to our letter and a summary of commissioned reviewers' criticism. An accompanying editorial published<sup>3</sup> by the journal stated that "it is each author's responsibility to make sure that statistical procedures are correctly used and valid for the study submitted".

**Journals charge authors to correct others' mistakes.** For one article that we believed contained an invalidating error, our options were to post a comment in an online commenting system or pay a 'discounted' submission fee of US\$1,716. With another journal from the same publisher, the fee

## STATISTICAL ANALYSIS

### Three common errors



As the influential twentieth-century statistician Ronald Fisher (**pictured**) said: "To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of."

Too many of our post-publication reviews were indeed post mortems. Some studies used inappropriate or non-randomization methods, despite stating that their studies were randomized (see, for example, ref. 5 and [go.nature.com/x2l9zz](http://go.nature.com/x2l9zz)). Others described mathematically or physiologically impossible results: *p*-values greater than 1, or an average height change of about 7 centimetres in adults over 8 weeks<sup>4,6</sup>.

Frequent errors, once recognized, can be kept out of the literature with targeted education and policies. Three of the most common are outlined below. These and others are described in depth in an upcoming publication<sup>7</sup>.

#### 1 Mistaken design or analysis of cluster-randomized trials.

In these studies, all participants in a cluster (for example, a cage, school or hospital) are given the same treatment. The number of clusters (not just the number of individuals) must be incorporated into the analysis. Otherwise, results often seem, falsely, to be statistically significant<sup>8,9</sup>. Increasing the number of individuals within clusters can increase power, but the gains are minute compared with increasing clusters. Designs with only one cluster per treatment are not valid as randomized experiments, regardless of how many individuals are included.

#### 2 Miscalculation in meta-analyses.

Effect sizes are often miscalculated when meta-analysts are confronted with incomplete information and do not adapt appropriately. Another problem is confusion about how to calculate the variance of effects. Different study designs and meta-analyses require different approaches. Incorrect or inconsistent choices can change effect sizes, study weighting or the overall conclusions<sup>4</sup>.

#### 3 Inappropriate baseline comparisons.

In at least six articles, authors tested for changes from the baseline in separate groups; if one was significant and one not, the authors (wrongly) proposed a difference between groups. Rather than comparing 'differences in nominal significance' (the DINS error) differences between groups must be compared directly. For studies comparing two equal-sized groups, the DINS error can inflate the false-positive rate from 5% to as much as 50% (ref. 10).

## FIXING POST-PUBLICATION REVIEW

Publishers, editors and researchers must all up their game.

	How to prevent statistical errors in submissions	How to streamline post-publication corrections
Research teams	Tap statistical expertise in the design and analysis of studies from the start. Describe analyses thoroughly.	Curate data and computer code so that they can be made easily available (for a registry of public data repositories, see <a href="http://www.re3data.org">www.re3data.org</a> ).
Manuscript editors	Create protocols to identify papers that need statistical scrutiny and send them to qualified reviewers.	Address readers' concerns swiftly. Use formal expressions of concern as an alert that work is under scrutiny — rather than for condemnation.
Journals and publishers	Require raw data and analysis code to be made available during review.	Create protocols to manage expressions of concern. State clearly who readers should contact and train editors to navigate protocols. Waive publication fees and paywalls for expressions of concern and retractions.

was £1,470 (US\$2,100) to publish a letter. Letters from the journal advised that “we are unable to take editorial considerations into account when assessing waiver requests, only the author’s documented ability to pay”. The Committee on Publication Ethics, an independent body that provides advice on how to handle research misconduct, asserts that readers should not have to pay to read retractions. To our knowledge, no authority has discussed whether third parties should be charged to correct errors.

**No standard mechanism exists to request raw data.** When we were able to access data online, we could quickly confirm suspected errors. In at least two cases, we requested data from the authors but received summaries of calculations instead. Sometimes we received no data at all, at which point it was not clear whether journal staff should step in. One journal did retract a paper when its authors refused to show their data or explain discrepancies that we had identified and alerted the journal to in a letter<sup>4</sup>.

Working directly with authors can delay correction. After we contacted authors about another paper, they offered to reanalyse the data to address our concerns. After a month with no response, we submitted a letter of concern to the journal. The letter was peer-reviewed and accepted within three weeks. The authors, when made aware of the pending publication of our letter, e-mailed us to state that they would prepare a reply, and we asked the journal not to publish our letter so that we could collaborate with the original authors. That process is ongoing, ten months after we identified the error.

**Informal expressions of concern are overlooked.** Although online platforms such as PubMed Commons offer a convenient way to comment on published papers, they do not include a mediating role for journal editors, and the comments

are not incorporated into the literature. Posted concerns are rarely prominent on journals’ websites and are not cross-referenced in any useful way. As a result, readers may assume that a flawed paper is correct, potentially leading to misinformed decisions in science, patient care and public policy.

In one case, we chose to post a comment on the journal website and on PubMed Commons after months of private correspondence, in which the authors shared some supplementary data and said that they were preparing a full response. The concerns have been acknowledged but remain unaddressed 15 months after we contacted authors and the journal, and 6 months after we posted our comment (see [go.nature.com/fv8tr2](http://go.nature.com/fv8tr2)).

### WHAT CAN BE DONE?

Journals have guidelines for paper submissions and peer review. The Committee on Publication Ethics has outlined recommendations for journals to address problems in areas such as authorship and review. But there is little formal guidance for post-publication corrections. (For our recommendations, see ‘Fixing post-publication peer review’.)

Journals, publishers and scientific societies should standardize, streamline and publicize these processes. Authors and journals should share data and code quickly when questions arise. Researchers can aid this process by accessing statistical expertise for experimental design and analysis.

Ideally, anyone who detects a potential problem with a study will engage, whether by writing to authors and editors or by commenting online, and will do so in a collegial way. Scientists who engage in post-publication review often do so out of

a sense of duty to their community, but this important work does not come with the same prestige as other scientific endeavours. Recognizing and incentivizing such activities could go a long way to cleaning up the literature.

Our work was not a systematic search; we simply looked more closely at papers that caught our eye and that we were prepared to assess. We do not know the rate of errors or the motivations behind them (that is, whether they are honest mistakes or a ‘sleight of statistics’). But we showed that a small team of investigators with expertise in statistics and experimental design could find dozens of problematic papers while keeping abreast of the literature. Most were detected simply by reading the paper.

A more formal survey would help to determine whether our experiences reflect science in general and whether our recommendations are feasible or effective. Others working to correct the scientific record have encountered similar challenges. Ben Goldacre, a physician and campaigner who is leading COMPare, a project that checks that clinical trials report the outcomes they said they would, told Retraction Watch: “This is a phenomenally laborious process. Not a week goes by that we don’t curse the day we set out to do this.”

Robust science needs robust corrections. It is time to make the process less onerous. ■

**David B. Allison** is a distinguished professor in the Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Alabama, USA.

**Andrew W. Brown** is a scientist in the Office of Energetics and the Nutrition Obesity Research Center, University of Alabama at Birmingham, Alabama, USA.

**Brandon J. George** is a statistician in the Office of Energetics, University of Alabama at Birmingham, Alabama, USA.

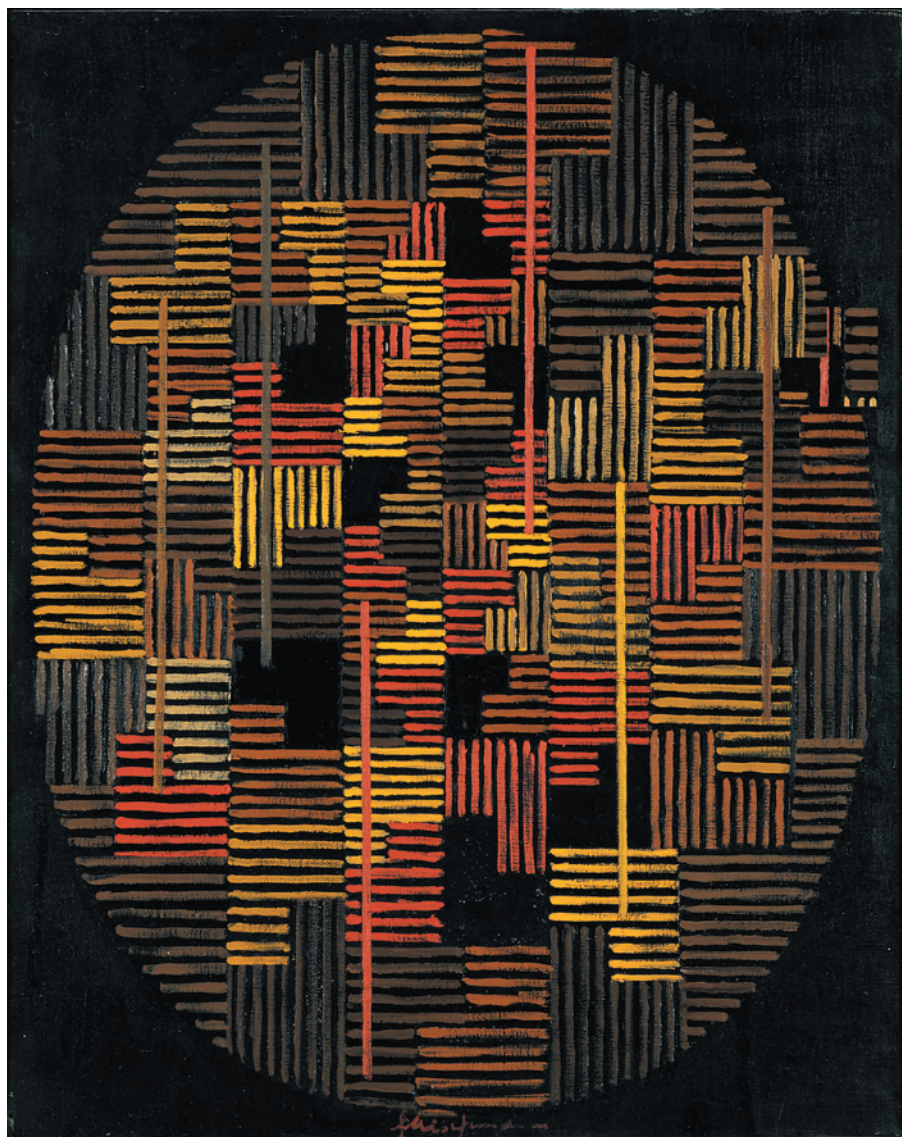
**Kathryn A. Kaiser** is an instructor in the Office of Energetics and the Nutrition Obesity Research Center, University of Alabama at Birmingham, Alabama, USA. e-mail: [dallison@uab.edu](mailto:dallison@uab.edu)

1. Brown, A. W. *et al.* *Child. Obes.* **10**, 542–545 (2014).
2. Li, P. *et al.* *Obes. Facts* **8**, 127–129 (2015).
3. Hauner, H. *Obes. Facts* **8**, 125–126 (2015).
4. George, B. J., Brown, A. W. & Allison, D. B. *J. Paramedical Sci.* **6**, 153–154 (2015).
5. George, B. J., Goldsby, T. U., Brown, A. W., Li, P. & Allison, D. B. *Int. J. Yoga* **9**, 87–88 (2016).
6. Thomas, D. M. *et al.* *World J. Acupunct. Moxibustion* **25**, 66–67 (2015).
7. George, B. J. *et al.* *Obesity* (in the press).
8. Brown, A. W. *et al.* *Am. J. Clin. Nutr.* **102**, 241–248 (2015).
9. *Obesity* **23**, 2522 (2015).
10. Bland, J. M. & Altman, D. G. *Am. J. Clin. Nutr.* **102**, 991–994 (2015).

D.B.A. declares competing financial interests: see [go.nature.com/hshkkk](http://go.nature.com/hshkkk) for details.

**“Authors and journals should share data and code quickly when questions arise.”**





Elements of Adolf Fleischmann's 1956 *Composition 71 N.Y.* recall his early work in medical imaging.

## IMAGING

# Medical modernist

**Thomas Schnalke** extols the dual genius of pathology sculptor and abstract artist Adolf Fleischmann.

**H**e created stunningly lifelike moulages: wax models of the human body in states of disease, once used for training doctors. Later, he turned to art, rising in the 1950s to become a star of the US abstract scene with paintings featuring vibrantly hued geometric shapes. Adolf Fleischmann (1892–1968) had impacts on medicine and art that were equally powerful and strangely divided.

This year, two Berlin exhibitions (for which I have contributed to the catalogues)

will explore Fleischmann's oeuvre: *Surfaces* at the Berlin Museum of Medical History at the Charité and the Adolf Fleischmann Retrospective at Daimler Contemporary Berlin. Whereas *Surfaces* is a survey of Fleischmann's varied life, focusing on medical works made between 1917 and 1927, the Daimler Contemporary retrospective concentrates on Fleischmann's artistic career in the United States, between 1952 and 1965.

The German-born Fleischmann trained

## **Surfaces: Adolf Fleischmann — Crossover Between Art and Medicine**

Berlin Museum of Medical History at the Charité.  
28 April – 11 September 2016.

## **Adolf Fleischmann Retrospective**

Daimler Contemporary Berlin.  
30 April – 6 November 2016.

as a graphic illustrator, then studied fine art in Stuttgart from 1911 to 1913. Heavily wounded in the First World War, he moved to neutral Switzerland in 1917 to work as a medical sculptor at Zurich's Surgical University Clinic. With the encouragement of the eminent moulage-maker Luise (Lotte) Volger and under clinic head and eminent surgeon Paul Clairmont, Fleischmann built up a unique collection of 400 surgical moulages over 10 years. These documented, in graphic 3D, trauma, pathological changes in the body, and therapeutic interventions visible on the patient's skin — such as wounds caused by strong electrical currents, swellings of the thyroid gland and side-effects of X-rays, such as skin atrophy. Impressive moulages of this kind will be on display at the Museum of Medical History.

Medical moulage-making, which had begun in cities including Jena, Germany, in the early nineteenth century, blossomed from the 1850s in the European medical centres of London, Paris, Vienna and Berlin. Around 1900, it spread around the world, coexisting with photography and other forms of graphic medical illustration until the 1950s, when the colour slide finally reached a satisfactory technical standard.

Creating a moulage involved taking a plaster cast of an area of the patient's skin and filling it with coloured liquid wax. Once detached, the wax shell was painted and finished from life to capture every nuance of form and colour, creating a perfect illusion for teaching. Although the process was clearly mimetic, the observational skill demanded was superb training for the artist's eye: in the topography of diseased and traumatized skin, Fleischmann could study organic form and detect graphic patterns and gradations of colour.

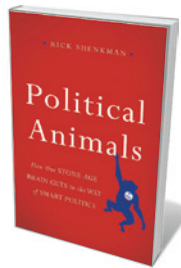
Although Fleischmann's moulages are unsigned, he did sign other works in his medical oeuvre, indicating that he felt they stood out visually and even artistically. These are 30 histopathological drawings of skin tissue, held at the Zurich Moulage Museum and largely overlooked. In Zurich between 1918 and 1927, Fleischmann used a microscope to make unprecedentedly subtle and accurate ink drawings of the dermatological evidence of diseases, such as the scaly skin disorder ichthyosis vulgaris, Hodgkin's lymphoma or the systemic autoimmune condition lupus erythematosus. He documented the intricate details of complex

structures and interactions of cells, nerves and veins, building his scientific understanding of visible organization and structure. As with the moulages, he reproduced form and colour; but with the drawings he also did more. There is a dynamical element in his mastery of line: the illustrations reveal a subtle movement, sublime gestures, the hidden contours under the skin. In these images, Fleischmann liberates himself as an artist.

He had been striving to become a fine artist, and to be seen as one, since the 1920s in Zurich. Here, he was able to absorb expressionist and cubist artworks, in particular those of Munich's Blue Rider group, which included Wassily Kandinsky and Franz Marc. But his path to acceptance was long and strewn with obstacles. The rise of Nazism and the Second World War forced moves to France, Spain, Italy and, post-war, back to Paris. Studying the work of artistic luminaries Robert Delaunay and Piet Mondrian, he finally arrived at his own distinct style in 1950. Settling in New York City two years later, he produced a stream of outstanding abstract paintings and prints that drew heavily on the urban elements of his new home.

Fleischmann only occasionally returned to medical imaging. However, there are hints of his microscopic drawings in several of

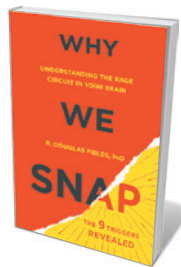
## Books in brief



### Political Animals: How Our Stone-Age Brain Gets in the Way of Smart Politics

Rick Shenkman BASIC (2016)

Politics often seems an arena for the irrational, in which drought can affect voting and candidates' egregious faults are ignored by die-hard fans. Journalist Rick Shenkman sees the cause as a poor fit between our Palaeolithic brains and today's knotted complexities. He liberally draws on psychology (from the likes of Daniel Kahneman) and political science to isolate four key failings among voters, including inept 'readings' of politicians. If democratic reform is to succeed, he argues, we must begin with self-reform.



### Why We Snap: Understanding the Rage Circuit in Your Brain

R. Douglas Fields DUTTON (2016)

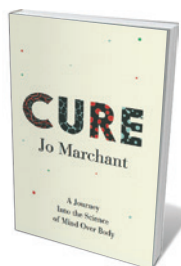
The tug of a pickpocket lifting his wallet spurred neurobiologist Douglas Fields to pin the man to the pavement — then motivated him to decode the brain's "rage circuitry". Synthesizing his own and others' research and scores of case studies, Fields argues that many apparently inexplicable cases of violent rage are down to a clash between hard-wiring in the brain's hypothalamus, amygdala and limbic system, and nine rage triggers, from life-or-death situations to threats to social order. He shows, too, how factors such as chronic stress can lower that flashpoint. Cogent and timely.



### Menagerie: The History of Exotic Animals in England

Caroline Grigson OXFORD UNIVERSITY PRESS (2016)

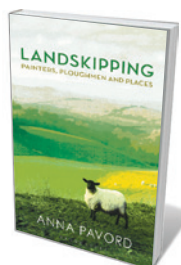
Hyenas roamed medieval Oxfordshire, reveals zoologist Caroline Grigson in this incisive chronicle of exotic visitations to England's shores. These and other animals in Henry I's park were a precursor to the Tower of London menagerie, which by the thirteenth century boasted a polar bear that swam in the Thames. The acquisition of wild beasts, initially the whim of status-hungry monarchs and a by-product of exploration, later became a public obsession and a focus for natural historians such as Hans Sloane — while the flood of monkeys and apes prompted early stirrings of evolutionary thought.



### Cure: A Journey into the Science of Mind Over Body

Jo Marchant CROWN (2016)

Science writer Jo Marchant probes the impact of mental states on physical health in this well-researched study of "mind-body medicine". There is much compelling science here, such as the emerging field of psychoneuroimmunology; and reminders of the negative effects of stress and poverty on health are salutary. However, the attention paid to often expensive alternative practices acknowledged by Marchant as unscientific sits oddly next to less controversial techniques — and the documented (and cheaper) benefits for low mood of a walk in the woods fail to feature.



### Landskipping: Painters, Ploughmen and Places

Anna Pavord BLOOMSBURY (2016)

The bosky glories of the British landscape were 'born', culturally, in the eighteenth century, and soon celebrated by luminaries such as Thomas Gainsborough and William Wordsworth. Anna Pavord traces their paths, skipping across scapes and interweaving a supple narrative of her own experience of place from Wales to Cumbria. Adding earthiness to lyricism are passages by agricultural writers such as William Cobbett of *Rural Rides* (1830), who abhorred the enclosure of common land, and Arthur Young, who supported it. [Barbara Kiser](#)

MOULAGE MUS. OF ZÜRICH UNIV. AND UNIV. HOSPITAL/PHOTO: BERNHARD STRAUSS



A moulage depicting skin damage from X-rays.

his late works of art, such as the oil painting *Composition 71 N.Y.* (1956), highlighted in the Daimler Contemporary show. Rounded motifs call to mind a microscope lens, for instance, and intricate grids of horizontal and vertical lines — combined with the 'flickering' appearance of the colours — give the pieces a visual dynamic. Perhaps this is why Fleischmann's artworks, although radically abstract in composition, appear so breathtakingly lively. ■

**Thomas Schnalke** is director of the Berlin Museum of Medical History at the Charité. e-mail: [thomas.schnalke@charite.de](mailto:thomas.schnalke@charite.de)



# The Crick Institute unpeeled

Ewen Callaway finds smart design fostering collaboration at London's biology super-lab.

London is having a basement boom. With property at a premium and height restrictions in place across most of the city, well-heeled homeowners in need of an extra bedroom or a bowling alley are burrowing underground. King's Cross, the city's northern transport hub, boasts the cellar of all cellars. Larger than a football pitch and 16 metres deep in places, it holds not a Russian oligarch's subterranean swimming pool, but the bio-containment laboratories and most sensitive scientific equipment of the Francis Crick Institute, London's new biology super-lab (see *Nature* 522, 406–408; 2015).

The first of the roughly 1,600 scientists and other staff members who will work at the Crick are set to move in later this year; the building will reach capacity by 2021. The Crick replaces three ageing laboratories in and near London. Its public and private funders, including Cancer Research UK and the Wellcome Trust, both in London, hope that it will become a beacon for UK life sciences.

Not long ago the Crick was, far from a beacon, one of London's most expensive holes in the ground, excavated from a plot hemmed in by the British Library, St Pancras International station and attendant railway, and an underground gas main. "We had to go down," explains Robert Partridge, a director at AKT II, the London-based firm in charge of the building's structural engineering. The whole 93,000-square-metre edifice is stabilized and isolated from vibrations by 14-metre-tall, 15-tonne cement and stainless-steel columns (magnetic materials are a no-no around many scientific instruments). Researchers can thank these columns when their electron microscopes remain unaffected by the rumbling of nearby train and tube lines, the bustle of central London and even the movements of the building's workforce (see *Nature* 518, 464–465; 2015).

The Crick's two-storey steel roof is another marvel of engineering (not to mention design — its undulations resemble an overturned ship with its hull split in two). It houses all the building's heating and ventilation infrastructure, including an air duct large enough to hold one of the capital's red double-decker buses. And it supports one of the biggest



The Crick will boast a public-engagement auditorium.

solar-panel arrays in London. The building is designed to cut back on energy and water use: for example, energy created as a by-product of powering the building is also used to heat it. And its 'brown roof' of native plants, like a green roof, helps to insulate the building and provides habitats for wildlife.

Like many twenty-first-century labs, the Crick is open-plan. Few walls separate lab benches in each of its four quadrants, and members of different groups work in view of one another, if not shoulder to shoulder. In theory, these spaces should facilitate collaboration and allow flexibility as groups swell, shrink and shift their focus. Whether this works or not is up for debate. In other open labs, I have seen scientists erect their own ad hoc barriers out of lab manuals or pipette racks to achieve a level of privacy.

➔ **NATURE.COM**  
For more on science  
in culture see:  
[nature.com/  
booksandarts](http://nature.com/booksandarts)

This quest for openness and collaboration goes even further at the Crick. Group meeting areas, tea-and-coffee-making facilities and

even solitary study desks are located in the centre of each floor. Glass walls and an open atrium make it possible to see colleagues on different floors across the building. Even the cafeteria is intentionally small to foster chance encounters. The Crick is designed to bring researchers together, even if they actually want to get away.

The building's architects, HOK, were also tasked with showing the Crick researchers to the public. Passers-by will be able to see into the building, and those with half an hour to spare are encouraged to pop into the public exhibition and engagement spaces just inside its cathedral-like entrance.

These design elements are a physical manifestation of the institute's mantra, "discovery without boundaries". It will be years before we know whether the institute has achieved goals such as attracting the world's most talented early-career scientists to Britain, or helping to catalyse a robust life-sciences economy. And we will probably never know whether the building's design has had a role in meeting these lofty ambitions. The Crick's success as a lab will rely just as heavily on nuts-and-bolts engineering as it will on the grander and more ambitious elements of its design.

In this regard, the Ray and Maria Stata Center (also known as Building 32) at the Massachusetts Institute of Technology in Cambridge offers a cautionary tale. Opened in 2004 to glowing architectural reviews, and also intended to catalyse interdisciplinary creativity, the Frank Gehry-designed building had a mixed reception from researchers. Some loved the new connections; some longed for a cave in which to concentrate. When the building developed leaks and cracks, the university sued the architects.

When the Crick fully opens later this year, it is possible that the workforce will not pay much attention to the building's behemoth of a basement or its hardworking roof. But somehow, I doubt it — not least because, in a building of engineering marvels named after one of the scientists who parsed the architecture of DNA, structure is likely to be much on their minds. ■

*Ewen Callaway is writes for Nature from London.*

MATT THOMAS/FRANCIS CRICK INST.

# Correspondence

## Mauritius culls threatened fruit bats

Mauritius has culled at least 20,000 flying foxes (*Pteropus niger*), the island's last surviving native species of fruit bat, in an attempt to reduce damage to lychee and mango crops.

Ironically, this cull was justified on the basis of a law enacted in late 2015 to improve protection of the country's rich biodiversity.

*Pteropus niger* is classified by the International Union for Conservation of Nature as threatened with extinction at the 'Vulnerable' level, and was protected under the previous law (F. B. V. Florens *Science* **336**, 1102; 2012).

The 2015 law makes provision for a special technical committee that is tasked with advising the government on the necessity of culling a species in the "national interest". But the decision to cull was announced in parliament before the new bill was enacted, and before the committee could be convened. The cull began within days of the law taking effect.

The biodiversity of Mauritius is among the most threatened worldwide. Consumers need to make an informed choice about supporting its fruit industry, which, in my view, shows disregard for the environment and international conventions. **F. B. Vincent Florens** *University of Mauritius, Réduit, Mauritius.* [vin.florens@uom.ac.mu](mailto:vin.florens@uom.ac.mu)

## Edited plants should not be patented

The development of genetically edited crops and plants raises another conundrum — that of intellectual-property protection (see *Nature* **528**, 319–320 (2015) and *Nature* **529**, 265 (2016)). We suggest that, in the interests of all parties, knowledge of these technologies should be made public without prior patenting.

Proponents of gene-editing technologies cannot have it both

ways. They hope that edited plants will not be classified as genetically modified organisms, to help to allay public fears. But then the plants would fall under the protection of the International Union for the Protection of New Varieties of Plants (UPOV), which excludes patenting.

A report from the scientific council of the French National Institute for Agricultural Research (INRA) argues that plants bred by any method should not be patented (see [go.nature.com/sp15nf](http://go.nature.com/sp15nf); in French). This recommendation is endorsed by the French Senate. Patents are granted for invention, not for having invested in the discovery or manipulation of native genes and sequences.

**John R. Porter** *University of Copenhagen, Denmark.*

**Jean-Louis Durand** *INRA URP3F, Lusignan, France.*

**Taline Elmayan** *INRA Centre Versailles-Grignon, France.* [jrp@plen.ku.dk](mailto:jrp@plen.ku.dk)

## Tibetan wildlife hemmed in

The Tibetan Plateau is a hotspot for biodiversity, roughly one-third of which is protected. Yet wildlife conservation efforts there are being disrupted by fences used to manage animal ranges, and by the expansion of roads and railways.

These projects threaten nature conservation and ecosystem health in the region. Affected animals include wild migratory herbivores such as the Tibetan antelope (*Pantholops hodgsonii*) and Przewalski's gazelle (*Procapra przewalskii*), both of which are endangered and need large areas of land for grazing and to promote gene flow between populations.

The plateau already has 51,300 kilometres of roads. More are planned, along with extensive railways to connect international borders (see [go.nature.com/whuz3z](http://go.nature.com/whuz3z)).

The fences are used to define farm boundaries and to aid rotational farming to reduce grassland degradation (see, for example, Z. Q. You *et al.* *Chinese Sci. Bull.* **58**, 2262–2268; 2013). Conservation authorities lack the power to override this government policy, which is backed by subsidies that reimburse local households for fence construction.

Policymakers must collaborate to avert the mounting threat to species and their trophic networks in these unique and important alpine ecosystems.

**Weihong Ji, Achyut Aryal** *Massey University, Auckland, New Zealand; and Gansu Agricultural University, China.*

**Junhu Su** *Gansu Agricultural University, China.*

[j.j.weihong@massey.ac.nz](mailto:j.j.weihong@massey.ac.nz)

## Balance research funds across Europe

We call for the European Union to push governments into keeping their research funding above subsistence level. This will ensure that scientists from across Europe can compete for Horizon 2020 research funding, not just those from the United Kingdom, Germany and Scandinavia.

Europe's research money is divided between the European Commission and national governments. The commission funds large, transnational collaborative networks in mostly applied areas of research, and the governments support small-scale, bottom-up science and their own strategic research programmes.

Some member states are not keeping their part of the bargain. Italy, for example, seriously neglects its research base. The Italian National Research Council has not overseen basic research for decades, being itself starved of resources. University funding has dwindled to a bare minimum. The ministerial initiative known as PRIN (Research Projects

of National Interest) has been defunct since 2012, apart from a few limited programmes for young researchers.

This year's PRIN allocation of a €92-million (US\$100-million) funding call to cover all research areas is too little, too late. Compare this with the annual French National Research Agency's allocation of up to €1 billion, or with Italy's €900-million annual contribution to the EU Seventh Framework Programme that ran in 2007–13. That resulted in a net annual loss of €300 million for Italian science.

To prevent distorted development in research among EU countries, national policies must be coherent and guarantee a balanced use of resources.

**Giorgio Parisi\*** *University of Rome, Italy.*

[giorgio.parusi@gmail.com](mailto:giorgio.parusi@gmail.com)

\*Supported by 69 signatories (listed at [go.nature.com/ye2iic](http://go.nature.com/ye2iic)).

## A toast to the deep-drilling idea

In your report on the revived drilling project to Earth's mantle, you mention a prominent group of National Academy of Sciences members at the Scripps Institution of Oceanography in California, who met in 1957 to outline the original Project Mohole (*Nature* **528**, 16–17; 2015). The year before, I had already put the case for deep drilling through the oceanic crust to reach the Mohorovičić discontinuity (see *Science* **124**, 686; 1956).

The members described their independent conception as happening over a "wine breakfast" by the ocean. This somewhat louche setting could not be a classed as a cocktail party, as you would have it. That might have been more accurate in my case, when a proper dry Martini at sunset may well have been involved.

**Frank B. Estabrook** *Jet Propulsion Laboratory, California, USA.* [festabrook@earthlink.net](mailto:festabrook@earthlink.net)



## ELECTROCHEMISTRY

# Photocatalysts in close-up

The water-splitting reaction is a promising route to renewable energy. Catalytic hotspots, and the best sites for co-catalyst placement, have now been pinpointed in a water-splitting catalyst, guiding future catalyst design. [SEE LETTER P.77](#)

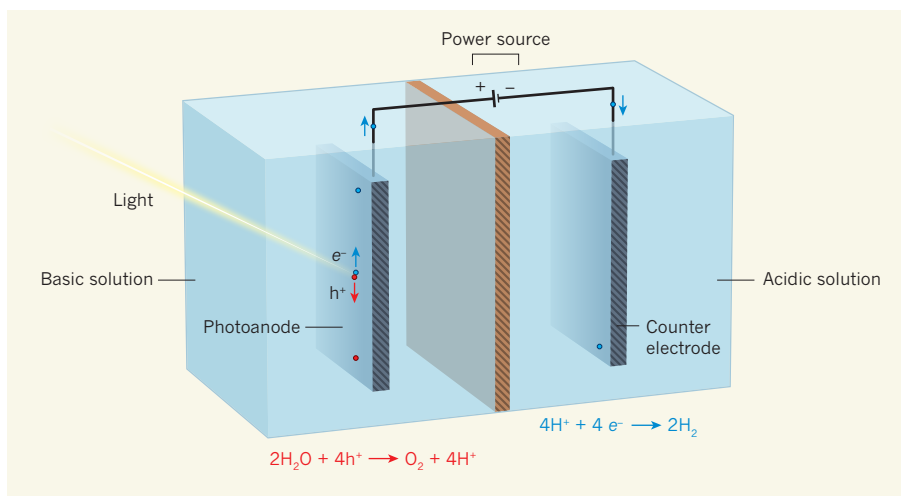
JOHAN HOFKENS  
& MAARTEN B. J. ROEFFAERS

Member states of the United Nations have now agreed to substantially reduce global emissions of carbon dioxide. This commitment challenges scientists to improve methods for producing renewable energy. An attractive approach pioneered<sup>1</sup> in 1972 is to produce hydrogen gas — a potential fuel — by ‘splitting’ water using solar energy and a solid photocatalyst, which mediates light-induced reactions. But no process has been commercialized because the conversion of light to fuel remains inefficient, despite more than 40 years of intense research. On page 77 of this issue, Sambur *et al.*<sup>2</sup> report exciting insights into how photocatalyst particles function at the molecular level. Their findings challenge current views of photocatalyst design, and bring the rational engineering of photocatalysts for splitting water a step closer.

Solar water splitting uses the energy contained in sunlight’s photons to generate hydrogen and oxygen gas from water. Although many renewable energy sources are known, and will be needed to satisfy global energy demands, hydrogen has the advantage of being a clean fuel, because only pure water is formed during its combustion. Direct water splitting requires highly energetic ultraviolet light, but this is efficiently filtered out by Earth’s atmosphere. Photocatalysts can induce this reaction using lower-energy ultraviolet photons, and even, in some cases, by using photons of visible light.

Solid photocatalysts capture photons and temporarily transfer photonic energy to an electron, exciting it into the catalyst’s conduction band and leaving behind a ‘hole’ in the valence band that behaves as a positively charged particle. The resulting electron–hole pair can induce chemical reactions at the photocatalyst’s surface. For example, the holes can oxidize water to form oxygen gas and protons ( $H^+$  ions), and the negatively charged electrons can reduce these protons to form hydrogen gas. Photocatalytic ‘cascades’ of this sort can therefore be used to store solar energy in chemicals such as hydrogen.

Electrons can also recombine with holes to



**Figure 1 | A water-splitting cell.** The light-induced water-splitting reaction can be used to generate oxygen and hydrogen gases from water. In this set-up, ultraviolet light hits an electrode (the photoanode, made from a photocatalytic material such as titanium oxide,  $TiO_2$ ), generating negatively charged electrons ( $e^-$ ) and positively charged ‘holes’ ( $h^+$ ). The photoanode is immersed in a solution of a base in water, and is connected to an external electric power source. This source rapidly drains away the excited electrons and directs them towards a counter electrode immersed in an acidic solution (a source of protons,  $H^+$ ), where they form hydrogen gas. The holes remaining at the photoanode react with water to form oxygen. Sambur *et al.*<sup>2</sup> have studied the oxygen-forming reaction on titanium oxide at the nanoscale.

produce heat. This process should be avoided, however, because it competes with photocatalytic reactions and lowers the efficiency with which energy can be stored in chemicals. In practice, photocatalysts are often connected to an external electric power source that rapidly drains away the excited electrons; the withdrawn electrons are directed towards a separate catalyst to form hydrogen gas from protons (Fig. 1). The holes remaining at the photocatalyst (which, in this set-up, is called the photoanode) are thus more likely to oxidize water than if the electrons had not been withdrawn. An additional oxygen-evolution co-catalyst deposited on the photoanode can be used to further enhance water oxidation and to minimize unwanted electron–hole recombination.

Minimizing the amounts of photocatalysts and co-catalysts needed to fabricate a photoanode is essential, not only to reduce costs, but also because co-catalysts compete for photons by scattering or absorbing them<sup>3</sup>. Attempts to optimize photoanode designs are often based on the interpretation of studies involving large

amounts of material, in which underlying molecular processes are masked because their effects average out. Such approaches overlook inherent differences between the individual particles that make up the bulk material. Solid catalytic materials are structurally complex at the smallest scales — even seemingly perfect crystals contain atomic-scale imperfections at surfaces at which the reactions occur<sup>4</sup>.

Sambur *et al.* used two approaches to directly measure the performance of titanium oxide ( $TiO_2$ ) photocatalysts. First, they mapped out the locations of single oxidation and reduction reactions at the catalyst surface with nanometre-scale precision, by using fluorogenic reactants (precursors of fluorescent products)<sup>5–7</sup>. These measurements revealed hotspots for oxidation and reduction reactions, which, surprisingly, overlap by about 40 nm.

Second, the authors analysed the number of electrons that could be withdrawn from the photoanode as different parts of it were selectively illuminated. From these measurements, they deduced the efficiency

with which an absorbed photon at a certain location gives rise to an oxidation reaction or to an unwanted electron–hole recombination. They observed that the oxidation and reduction hotspots are also the most active recombination regions.

With information in hand about the variation of reactivity with position, Sambur and co-workers went on to selectively deposit oxygen-evolution co-catalysts at hotspots or at regions with intrinsically low activity. Unexpectedly, the authors found that oxygen evolution was most effectively improved when the co-catalyst was located at positions that had intrinsically low activity. These regions are currently not targeted by researchers because conventional deposition methods tend to place co-catalysts at more-active positions<sup>8</sup>.

Sambur and colleagues' work paves the way for the rational engineering of catalysts and for improved photoanode operation — the ability to map the nanoscale electrochemical efficiency and redox reactivity of photocatalysts is a major step forward. However, challenges remain before these findings can lead to the development of improved technology. For example, how can spatially selective co-catalyst deposition be easily and cost-effectively scaled up from the single-particle approach used by the authors? Furthermore, the spatial resolution of the experiments is limited to a few tens of nanometres. This means that the data still represent average performances for thousands of surface atoms, many of which are possible catalytic sites. There is ample room for improvement, for example by enhancing the resolution of the techniques used.

More broadly, the authors' insights into the functioning of titanium oxide photocatalysts demonstrate yet again the complexity of catalytic processes on solid materials in general<sup>9</sup>. Their breakthrough experiments underline how difficult it is for bulk-scale studies to correctly capture this complexity and to formulate guidelines for rational catalyst design<sup>10</sup>. ■

**Johan Hofkens** is in the Department of Chemistry, and **Maarten B. J. Roelfaers** is in the Department of Microbial and Molecular Systems, KU Leuven University, B-3001 Heverlee, Belgium.  
e-mails: johan.hofkens@chem.kuleuven.be; maarten.roelfaers@biw.kuleuven.be

1. Fujishima, A. & Honda, K. *Nature* **238**, 37–38 (1972).
2. Sambur, J. B. *et al.* *Nature* **530**, 77–80 (2016).
3. Gamelin, D. R. *Nature Chem.* **4**, 965–967 (2012).
4. Roelfaers, M. B. J. *et al.* *Nature* **439**, 572–575 (2006).
5. Roelfaers, M. B. J. *et al.* *Angew. Chem. Int. Edn* **48**, 9285–9289 (2009).
6. Zhou, X. *Nature Nanotechnol.* **7**, 237–241 (2012).
7. Tachikawa, T., Yamashita, S. & Majima, T. *J. Am. Chem. Soc.* **133**, 7197–7204 (2011).
8. Kim, T. W. & Choi, K. S. *Science* **343**, 990–994 (2014).
9. Zecevic, J., Vanbutsele, G., de Jong, K. P. & Martens, J. A. *Nature* **528**, 245–248 (2015).
10. Plessers, E. *et al.* *ACS Catal.* **5**, 6690–6695 (2015).

## AGEING

# A stretch in time

**Plots of survival against time for nematode worms in different conditions can be superimposed by rescaling the time axis. This observation has far-reaching implications for our understanding of the nature of ageing. SEE LETTER P.103**

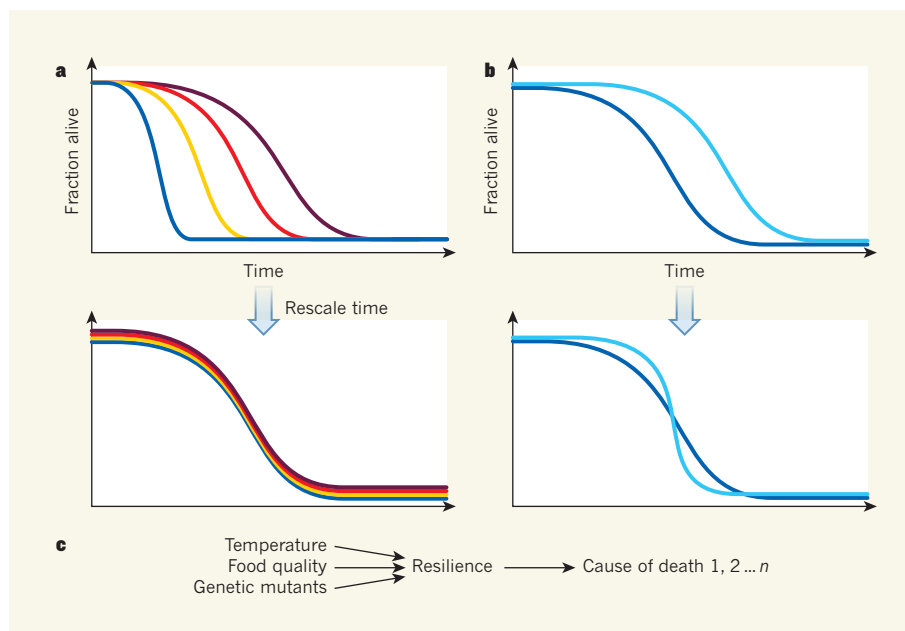
ZACHARY PINCUS

It is a long-standing mystery why, although the rate of ageing varies greatly among species, the effects of ageing are remarkably consistent. Closely related species such as mice and naked mole rats can have average lifespans that differ by more than tenfold, yet these species (and others as distant as yeast and humans) undergo similar molecular changes throughout ageing. In this issue, Stroustrup *et al.*<sup>1</sup> (page 103) rigorously demonstrate that the way in which the risk of death changes over the course of an organism's life is largely independent of the length of that life.

The nematode worm *Caenorhabditis elegans* is often used for studies of ageing, because it has a lifespan of about two weeks and is easy to cultivate in the laboratory. Stroustrup and colleagues used their previously developed<sup>2</sup>

'lifespan machine' to simultaneously measure the survival of tens of thousands of individual worms with 20-minute precision. The authors subjected the worms to a range of lifespan-altering conditions — different temperatures, a damaging compound, a lifespan-prolonging food source and several genetic mutations that extend or shorten lifespan. After analysing the lifespans of more than 100,000 individuals, they demonstrate that the overall shape of the survival curve (a plot of the fraction of the starting worm population alive at any given time) remains unchanged in different conditions. They find that these conditions act only to stretch or squeeze the curve along the time axis (Fig. 1a).

This is altogether unexpected, because it implies that the tested lifespan-altering manipulations change the probability of every possible cause of death in concert and to



**Figure 1 | Resilience as a measure of ageing.** **a**, Stroustrup *et al.*<sup>1</sup> measured the fraction of a population of nematode worms that remained alive over time in various conditions, constructing 'survival curves' (coloured lines) for each condition. The survival curves for many different conditions can be superimposed simply by rescaling the time axis. This observation implies that each condition alters the probability of every cause of death to the same extent. **b**, If this were not the case, the curves would not be superimposable. For example, if a fast-acting cause of death is made less likely, but not a slow-acting one, then the curve's shape will change in one particular region (as in the light blue compared to the dark blue curve). **c**, To obtain the scaling observed by Stroustrup and colleagues, there must be an organismal state, here dubbed 'resilience', that is influenced by many determinants of lifespan, and that is the sole determinant of the risk of death from any particular cause (different causes are represented by numbers).



exactly the same extent. To explain, because each cause of death plays out along a distinct timescale, if a particular condition were to decrease the odds of a fast-acting cause of death but not a slow-acting cause (or vice versa), a particular region of the overall survival curve would be altered, and thus its overall shape changed (Fig. 1b). Even interventions such as temperature shifts might be expected to affect different causes of death differently. After all, death is a biochemical process, and changing temperatures will alter the rates of different death-promoting chemical reactions differently, depending on the activation energies of those reactions. The authors' observation of almost-perfect timescaling across different conditions thus places specific constraints on how these conditions influence survival.

How can this surprising observation be explained? One possibility is that every cause of death in the worms has the same activation energy and responds identically to changes in food source, toxic exposures and diverse genetic mutations. Another is that worms have a single mechanistic cause of death.

A more plausible interpretation is that there is some intermediate state on which all the tested interventions converge, and which determines the risk of death from each possible cause (Fig. 1c). Could this intermediate state involve, for instance, the insulin/insulin-like growth factor signalling (IIS) pathway, which is central to many aspects of ageing across species<sup>3</sup>? No — the authors found that survival curves retain their shape even when the IIS pathway is inactivated, and in response to conditions and mutations known to act independently of the IIS pathway.

The most likely explanation for this intermediate state is that the risk of death is governed not by any single pathway, but by a property that arises from interactions between the various molecular processes that influence ageing. This property, perhaps best called 'resilience', would be an intrinsic biological property of ageing *C. elegans*, just as temperature and pressure are intrinsic thermodynamic properties of gases that emerge from the interactions of the constituent molecules.

The temperature of water in a whistling tea kettle provides an analogy for resilience. There are many ways in which to heat the water — on a stove, in a microwave, or even by adding a strong acid. However, whether a kettle whistles depends not on the source of the heat, but on the water temperature. Similarly, alterations in the molecular processes that contribute to resilience could change the rate of ageing (the heating rate of the water) without changing its underlying nature (the relationship between temperature and whistling).

The authors use detailed simulations to demonstrate how such a property could emerge. If resilience is a measure of the fraction of

biological processes in a densely interconnected network that have failed, then manipulations that alter a subset of these processes will extend or shorten lifespan without changing the shape of the survival curve. Alternatively, a single physical property, which is acted on by many molecular processes and affects the risk of death from diverse causes, could underlie an organism's resilience. Potential candidates for such physical properties include intracellular redox levels<sup>4,5</sup> or global protein solubility levels and turnover rates<sup>6,7</sup>. Whatever the case, the current work provides a strong constraint on any proposed molecular mechanism of resilience — measurements of the levels or activity of that mechanism must correlate exactly with lifespan across temperatures and among different genetic mutants.

This study suggests that concepts such as resilience and frailty, long used in the ageing literature, might have a concrete biological meaning. In particular, the Rockwood frailty index, which calculates the fraction of measured clinical markers considered to be in a deficient state<sup>8</sup>, is a close theoretical match for the authors' interpretation of resilience. Most importantly, these results demonstrate that, although students of ageing biology have learnt much about how to manipulate the rate of ageing, the nature of organismal frailty is almost completely unknown. The few interventions the authors identify that do change the shape of the survival curve in *C. elegans*

(such as a mutation that alters feeding ability, and another that alters function in mitochondria, the cell's energy centres) may point the way towards understanding this previously unappreciated biology.

Finally, although Stroustrup *et al.* consider only lifespan, increasing chronological lifespan does not necessarily increase the fraction of lifespan spent in good health<sup>9</sup>. Further work of a similar experimental and analytical rigour will be necessary to clarify the relationship between quality and quantity of life. ■

**Zachary Pincus** is in the Departments of Developmental Biology and Genetics, Washington University School of Medicine, St. Louis, Missouri 63110, USA.  
e-mail: zpincus@wustl.edu

1. Stroustrup, N. *et al.* *Nature* **530**, 103–107 (2016).
2. Stroustrup, N. *et al.* *Nature Methods* **10**, 665–670 (2013).
3. Kenyon, C. J. *Nature* **464**, 504–512 (2010).
4. Terman, A. & Brunk, U. T. *Antioxid. Redox Signal.* **8**, 197–204 (2006).
5. Kirstein, J. *et al.* *EMBO J.* **34**, 2334–2349 (2015).
6. Gelino, S. & Hansen, M. J. *Clin. Exp. Pathol. Suppl.* **4**, 006 (2012).
7. Vilchez, D., Simic, M. S. & Dillin, A. *Trends. Cell Biol.* **24**, 161–170 (2014).
8. Mitnitski, A. & Rockwood, K. *Biogerontology* <http://dx.doi.org/10.1007/s10522-015-9583-y> (2015).
9. Huang, C., Xiong, C. & Kornfeld, K. *Proc. Natl Acad. Sci. USA* **101**, 8084–8089 (2004).

This article was published online on 27 January 2016.

#### PARKINSON'S DISEASE

## Disorder in the court

**The native structure of the protein  $\alpha$ -synuclein, which is implicated in Parkinson's disease, is controversial. In-cell nuclear magnetic resonance now shows that it remains disordered when loaded into living cells. [SEE ARTICLE P.45](#)**

T. REID ALDERSON & AD BAX

The 3D structure of a biological molecule typically dictates its function. In an apparent paradox, intrinsically disordered proteins (IDPs) lack well-defined 3D structures, but this structural plasticity can confer diverse biological functions. The small protein  $\alpha$ -synuclein is an IDP *in vitro*, but whether it exists *in vivo* as a monomeric IDP<sup>1</sup> or as an ordered, helical tetramer<sup>2,3</sup> is a matter of dispute<sup>4</sup>. If  $\alpha$ -synuclein is an ordered tetramer *in vivo*, it might be possible to design drugs that stabilize this protein state and prevent the protein from forming toxic aggregates, which are associated with Parkinson's disease. On page 45 of this issue, Theillet *et al.*<sup>5</sup> describe a detailed analysis of the structure of  $\alpha$ -synuclein inside living cells.

Ever since the structure of DNA was

unveiled in 1953, the importance of understanding biology at the atomic level has been widely appreciated. But the standard method for resolving structures at atomic resolution, X-ray diffraction, relies on the generation of carefully grown crystals, which involves separating molecules of interest from their native environments. This process can bias the structures of molecules in ways that are hard to predict.

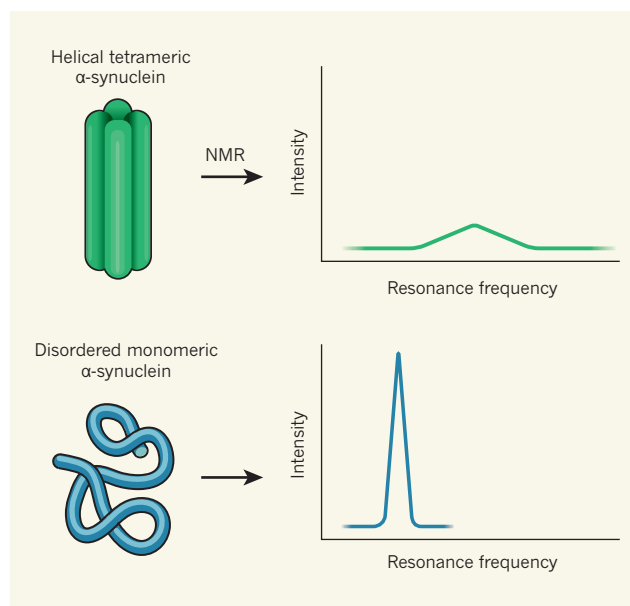
An alternative to X-ray diffraction is nuclear magnetic resonance (NMR) spectroscopy, which can be used to analyse molecules under the same conditions of pH, temperature and ionic concentration as when they are inside cells. In the past decade, it has become possible to use NMR spectroscopy to determine the structure and dynamics of proteins in living cells at atomic resolution<sup>6,7</sup>. In these in-cell studies, enrichment of molecules

of interest with the stable, non-perturbing isotopes nitrogen-15 or carbon-13 permits the NMR signals of these molecules to be filtered from the sea of other cellular components. In 2009, in-cell NMR was used to determine the first structure of a protein inside live bacteria<sup>6</sup>, but loading proteins of interest into mammalian cells in large-enough quantities and keeping these cells alive for long enough to perform NMR has been a challenge. Theillet *et al.* have overcome these obstacles, and have used in-cell NMR and a related technique, electron paramagnetic resonance (EPR) spectroscopy, to characterize the structure and dynamics of  $\alpha$ -synuclein inside mammalian cells.

$\alpha$ -Synuclein is abundant in many cell types, particularly neurons, where it is found in structures called presynaptic termini, which make contact with and signal to other cells<sup>8</sup>. The protein has a high affinity for certain types of membrane, and it is thought to modulate the release and recycling of presynaptic vesicles that contain neurotransmitter molecules<sup>9</sup>. Many biophysical studies have concluded that, in its natural state,  $\alpha$ -synuclein is an IDP that adopts  $\alpha$ -helical conformations when it binds to membranes, but these studies mainly used purification conditions that would have denatured (unfolded the structure of) the protein<sup>4</sup>. Thus, it could be that the protein's disorder in the absence of a membrane is a consequence of purification protocols, rather than an intrinsic property of the protein. By inference, this criticism could equally apply to many other IDPs.

In 2011, it was reported that, rather than being an IDP,  $\alpha$ -synuclein exists as a helical tetramer when purified from mammalian cells under non-denaturing conditions<sup>2</sup>. This finding took researchers by surprise and ignited a vigorous debate<sup>2,3,10</sup>. Proponents of the work argue that  $\alpha$ -synuclein forms tetramers in living cells, and suggest that these tetramers had dissociated during purification in previous studies<sup>2,3,10</sup>. They further propose that the tetrameric state, which might require a stabilizing cofactor *in vivo*<sup>10</sup>, acts as a storage mechanism that safeguards the protein from forming toxic aggregates<sup>2,3,10</sup>. However, subsequent attempts to confirm the existence of helical tetramers have failed<sup>1,4</sup>. Furthermore, other studies have shown that when  $\alpha$ -synuclein is added to *in vitro* samples in its monomeric, disordered state, it can aid the assembly of the presynaptic-vesicle fusion machinery<sup>9,11</sup>.

Theillet *et al.* introduced a bacterial form of  $\alpha$ -synuclein into five types of mammalian cell at concentrations close to those found *in vivo*.



**Figure 1 | Structure of a disordered protein in living cells.** There has been much debate about whether the protein  $\alpha$ -synuclein exists as a structured, helical tetramer in healthy cells, or whether it is an intrinsically disordered monomer with no fixed structure. Theillet *et al.*<sup>5</sup> used in-cell nuclear magnetic resonance (NMR) spectroscopy to address this question. If the protein formed a helical tetramer, then NMR spectroscopy would produce broad signals, which are characteristic of structured tetramers undergoing slow molecular reorientation. However, the authors find that the protein reorients rapidly, giving sharp NMR peaks at positions that are indicative of the disordered monomer.

They find compelling evidence that the protein remains highly disordered in these conditions, with each monomer changing shape rapidly. Remarkably, despite the fact that IDPs might be expected to be sensitive to degradation, bacterially produced  $\alpha$ -synuclein remained intact for days in the authors' cell lines — two of which are closely related to human neurons. Theillet and colleagues confirmed that the cells were not adversely affected by the  $\alpha$ -synuclein. Moreover, in a follow-up report published in *Nature Communications*<sup>12</sup>, the same group demonstrated that the cells retained the ability to repair damaged forms of bacterially produced  $\alpha$ -synuclein, indicating that the protein can interact functionally with the cells' molecular machinery.

By analysing the in-cell NMR and pulsed EPR spectra from  $\alpha$ -synuclein, Theillet and colleagues demonstrated that the signals from the in-cell protein were at similar positions to those of the disordered, *in vitro* reference protein — in particular, NMR signals were sharp and showed that the protein remains highly flexible and undergoes rapid molecular reorientation (Fig. 1). The atomic motions of the internalized  $\alpha$ -synuclein were slightly more sluggish than those of the reference protein, presumably because of the viscosity of the cellular environment, and the protein showed transient, localized hydrophobic and electrostatic interactions with components of the cytoplasm. The authors' analysis

provides the first insight into how individual residues in an IDP interact with the intracellular environment.

Finally, Theillet *et al.* showed that high intracellular concentrations of macromolecules and other cellular components can cause  $\alpha$ -synuclein monomers to become slightly more compact, perhaps protecting the monomers from forming the stable intermolecular interactions seen in the toxic, aggregated form of the protein.

Although this study demonstrates that  $\alpha$ -synuclein does not fold into a tetramer when introduced into mammalian cells, the authors' data do not fully exclude the existence of a tetrameric state. NMR signals from a small population of tetramers would be broad (an indication that a structure undergoes molecular reorientation only slowly) and weak (Fig. 1), and would therefore be invisible with in-cell NMR. Thus, the protein could exist in a dynamic equilibrium between monomeric and tetrameric states<sup>3,10</sup>. Up to 20% of  $\alpha$ -synuclein could exist in a tetrameric state at any one time, because the NMR signal intensities from the monomer accounted for 90%, plus or minus 10%, of the total amount of intro-

duced protein. Thus, the tetramer hypothesis could potentially be reconciled with Theillet and colleagues' results — but only if the intrinsic tetramer concentration is low, or if the otherwise intact machinery of the cells used in the work somehow lacks the ability to convert monomeric  $\alpha$ -synuclein into tetramers. ■

**T. Reid Alderson and Ad Bax** are in the Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. T.R.A. is also in the Department of Chemistry, University of Oxford, Oxford OX1 3TA, UK. e-mails: bax@nih.gov; reid.alderon@nih.gov

1. Fauvet, B. *et al.* *J. Biol. Chem.* **287**, 15345–15364 (2012).
2. Bartels, T., Choi, J. G. & Selkoe, D. J. *Nature* **477**, 107–110 (2011).
3. Dettmer, U. *et al.* *Nature Commun.* **6**, 7314 (2015).
4. Alderson, T. R. & Markley, J. L. *Intrins. Disord. Proteins* **1**, 18–39 (2013).
5. Theillet, F.-X. *et al.* *Nature* **530**, 45–50 (2016).
6. Sakakibara, D. *et al.* *Nature* **458**, 102–105 (2009).
7. Freedberg, D. I. & Selenko, P. *Annu. Rev. Biophys.* **43**, 171–192 (2014).
8. Lashuel, H. A., Overk, C. R., Oueslati, A. & Masliah, E. *Nature Rev. Neurosci.* **14**, 38–48 (2013).
9. Burré, J. *et al.* *Science* **329**, 1663–1667 (2010).
10. Luth, E. S., Bartels, T., Dettmer, U., Kim, N. C. & Selkoe, D. J. *Biochemistry* **54**, 279–292 (2015).
11. Burré, J., Sharma, M. & Südhof, T. C. *Proc. Natl Acad. Sci. USA* **111**, E4274–E4283 (2014).
12. Binolfi, A. *et al.* *Nature Commun.* <http://dx.doi.org/10.1038/ncomms10251> (2016).

This article was published online on 25 January 2016.



## OPTICAL PHYSICS

# Ultrashort light pulses shake atoms

The response of electrons in atoms to ultrashort optical light pulses has been probed by measuring the ultraviolet light emitted by the atoms. This reveals that a finite time delay occurs before the response. [SEE LETTER P.66](#)

KYUNG TAEC KIM

Characterization of the atomic and molecular response to a light wave is the first step in understanding the fundamental interactions between light and matter. Because the response of electrons in atoms to light waves is so rapid, tremendous efforts have been made to develop light sources that produce pulses as short as possible to study such ultrafast dynamics. In this issue, Hassan *et al.*<sup>1</sup> (page 66) report the generation of sub-femtosecond light pulses in the visible and near-visible spectral ranges (1 femtosecond is  $10^{-15}$  s). The authors used these ultrashort pulses to probe nonlinear responses of electrons in krypton atoms to light, and find that there is a finite response time.

Atoms are composed of negatively charged electrons and a positively charged nucleus. The response of an atom to a light wave is often considered to be akin to the behaviour of a system of two masses — one for a bound electron and the other for the rest of the atom — connected by a spring (Fig. 1). When the atom is exposed to the light wave, the electron moves with respect to the nucleus because of the Lorentz force (the combination of electric and magnetic forces that acts on a point charge in the presence of an electromagnetic field), resulting in a charge separation. In a weak light field, the proportionality between the field strength and the charge separation is linear, which means that the electron oscillates with the same frequency as that of the applied field.

But the situation becomes complicated when atoms are exposed to a strong field (but not strong enough to cause much ionization). Electron oscillation becomes nonlinear with respect to the applied light field as the electron moves farther from the nucleus. The oscillating electron can emit radiation with a broader spectrum than that of the applied light. This radiation offers a key to understanding the nonlinear response of the atom to the strong light field.

To investigate such nonlinear responses, Hassan and colleagues have generated optical attosecond pulses — bursts of light in the visible and near-visible spectral ranges that last for several hundreds of attoseconds (1 as is  $10^{-18}$  s). First, the authors passed laser pulses through a hollow fibre filled with

neon gas<sup>2</sup>, which broadened the spectrum of the pulses to encompass the near-infrared, visible and ultraviolet regions (equivalent to photon energies of between approximately 1.1 and 4.6 electronvolts). The spectrum was then divided into four different bands, each of which was optimized to produce shorter-duration pulses. Finally, the researchers combined the four pulses, controlling the pulses' phases (their progression through their electromagnetic wave cycles) in such a way as to generate optical attosecond pulses.

The resulting light-field synthesizer has an extraordinary performance: the initial laser-pulse duration of 22 fs is compressed to 975 as without serious energy loss. The synthesized pulse looks like a half-cycle pulse (Fig. 1). This is the first demonstration of the generation of isolated sub-femtosecond light pulses in the visible and ultraviolet spectral ranges.

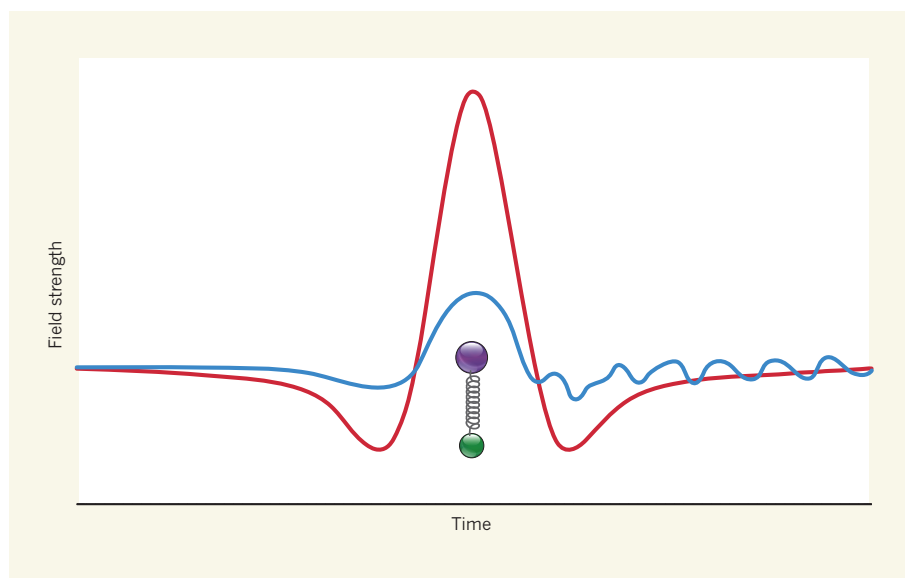
Hassan *et al.* went on to irradiate krypton atoms with intense optical attosecond

pulses. The temporal profile of these pulses changes dramatically depending on their phase. Accordingly, the ultraviolet spectra generated by the krypton atoms displayed pronounced phase-dependent amplitude modulation. The data reveal that the nonlinear response of bound krypton electrons to light is delayed, and that the delay increases, by up to 115 as, with the applied strength of the optical field.

The authors also devised and verified an equation to describe the nonlinear response, and showed that this reconstructs the experimental results, including the delay. They attribute the response to an increase in the total nonlinear polarizability of the atom during its interaction with the optical attosecond pulse.

This work has important implications for future research in ultrafast science. Until now, attosecond X-ray pulses have been the main tool for ultrafast 'pump-probe' experiments<sup>3</sup>, acting as both the stimulus for a physical effect and a means of probing its outcomes. However, the applications of these X-ray pulses are quite limited because of their weak intensity. By contrast, Hassan and colleagues' optical attosecond pulses are intense enough to initiate strong-field processes, opening up routes to new experiments. Moreover, such pulses could be applied to molecules and solids, which may produce different nonlinear responses depending on their proton dynamics<sup>4</sup> and structure<sup>5</sup>.

The demonstration of optical attosecond pulses and their application to probe non-instantaneous responses of atoms is



**Figure 1 | Bound electron motion in a half-cycle light wave.** An atom in a light wave can be considered to behave like a system of two masses — a light mass for a bound electron (green ball), and a heavy mass for the rest of the atom (purple ball) — connected by a spring. The electron oscillates with respect to the rest of the atom because of the electric force associated with the applied light field (red line indicates the light field; here, the light pulse corresponds to half an electromagnetic wave cycle). The oscillating electron can emit nonlinear radiation (blue line). Hassan *et al.*<sup>1</sup> have used sub-femtosecond light pulses in the visible and near-visible spectral ranges to probe the radiation emitted by krypton atoms.

remarkable. There is, however, one issue still to be resolved. The analysis reported in this work relies on measurements of spectral amplitude. To access the nonlinear response directly, both the amplitude and phase of radiation from atoms driven by a light field should be characterized. This is challenging, because there is no easy way to measure the phase of radiation

at ultraviolet wavelengths. If such methods are realized, they would open up yet another horizon in ultrafast science. ■

**Kyung Taec Kim** is at the Center for Relativistic Laser Science, Institute for Basic Science, and Department of Physics and Photon Science, Gwangju Institute of Science

and Technology, Gwangju, Republic of Korea.  
e-mail: kyungtaec@gist.ac.kr

1. Hassan, M. Th. *et al.* *Nature* **530**, 66–70 (2016).
2. Nisoli, M. *et al.* *Opt. Lett.* **22**, 522–524 (1997).
3. Corkum, P. B. & Krausz, F. *Nature Phys.* **3**, 381–387 (2007).
4. Baker, S. *et al.* *Science* **312**, 424–427 (2006).
5. Itatani, J. *et al.* *Nature* **432**, 867–871 (2004).

## CANCER

# Fibroblasts for all seasons

**Connective-tissue cells known as fibroblasts display an increasing spectrum of functions. Different fibroblast subtypes are now shown to either promote or suppress inflammation-associated intestinal cancers.**

ERWIN F. WAGNER

Cancer-associated fibroblasts (CAFs) are a key cell population in the tumour stroma, the term used for all cells of the tumour microenvironment except the cancerous ones. CAFs typically originate from mesenchymal cells, which are present in several healthy tissues. They frequently promote cancer progression by inducing cell proliferation, inflammation, blood-vessel growth and metastasis. However, they can also restrain tumour formation<sup>1</sup>. Two papers<sup>2,3</sup> in *The Journal*

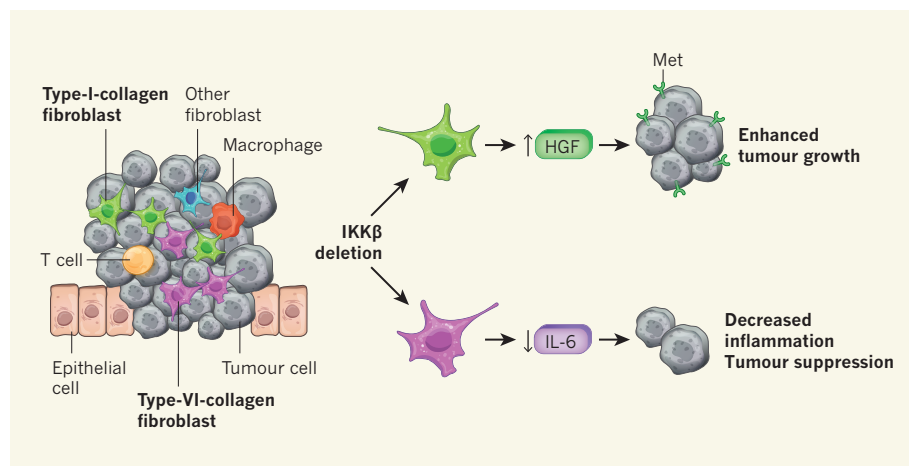
of *Experimental Medicine* highlight the complexity of these cells and remind us to be cautious in contemplating their use in therapeutic applications. Both research groups studied the effect of fibroblast-specific inhibition of the NF- $\kappa$ B–IKK signalling pathway, a major mediator of inflammation and cancer<sup>4</sup>, on inflammation-associated colorectal cancer. Pallangyo *et al.*<sup>2</sup> find that such inhibition promotes cancer development in mice, whereas Koliarakis *et al.*<sup>3</sup> report a suppressive effect.

Pallangyo and colleagues used the carcinogen azoxymethane (AOM) in conjunction

with the inflammatory agent dextran sodium sulfate (DSS) to induce colitis-associated cancer in mice. They inhibited NF- $\kappa$ B–IKK signalling in CAFs by specifically deleting the gene that encodes IKK $\beta$ , and found that this promoted proliferation of cancerous intestinal epithelial cells. It also suppressed tumour-cell death, induced the formation of blood vessels and enhanced the recruitment of immune cells, all features that contribute to enhanced tumour growth. Furthermore, IKK $\beta$ -deficient CAFs showed activated TGF- $\beta$  signalling, a pathway that can promote cell proliferation, and secreted elevated levels of hepatocyte growth factor (HGF), a major growth factor produced by CAFs (Fig. 1). The authors also show that pharmacological inhibition of Met, the receptor for HGF, reduced tumour growth in these mice.

Surprisingly, Koliarakis *et al.* come to the opposite conclusion, despite using similar protocols for inducing colitis-associated cancer and deleting IKK $\beta$ . The researchers report that inhibition of NF- $\kappa$ B–IKK signalling led to a reduction in the incidence and number of intestinal tumours. They observed reduced epithelial-cell proliferation and immune-cell infiltration, and lowered expression of inflammatory cytokine proteins, such as interleukin-6 (IL-6). A similar result was obtained when IKK $\beta$  expression was inhibited in the fibroblasts of mice that mimicked the familial colorectal cancer adenomatous polyposis (APC), but only when the mice were subjected to DSS-induced inflammation. This indicates that the tumour-suppressive effect of inactivating IKK $\beta$  in fibroblasts is restricted to cases of inflammation-associated colorectal cancer.

How can these results, which at first glance seem contradictory, be explained? One possibility lies in the fact that the studies use slightly different strategies to delete the gene that encodes IKK $\beta$  (technically speaking, they use different conditional alleles and different collagen gene promoters to express the Cre recombinase). The genetic background of the mice, the timing of IKK $\beta$  deletion and the population of fibroblasts targeted in the two experimental settings also differ (Fig. 1), as does the environment and possibly the resident microorganisms of the mutant mice. Pallangyo and colleagues' deletion of IKK $\beta$  involved treating mice with the molecule tamoxifen, and the deletion effectively started at the tumour-initiation stage. Koliarakis and colleagues used



**Figure 1 | Fibroblast functions in inflammation-associated colorectal cancer.** Tumours contain non-cancerous cells that can influence the growth and progression of the tumour. Pallangyo *et al.*<sup>2</sup> and Koliarakis *et al.*<sup>3</sup> studied the effect of loss of the signalling protein IKK $\beta$  in fibroblasts in mouse models of inflammation-associated colorectal cancer. When Pallangyo *et al.* deleted IKK $\beta$  at tumour initiation in type-I-collagen-producing fibroblasts, they observed enhanced tumour growth, seemingly mediated through fibroblast production of hepatocyte growth factor (HGF), which binds to the receptor Met on tumour cells. By contrast, when Koliarakis *et al.* constitutively deleted IKK $\beta$  in a more-restricted population of type-VI-collagen-producing fibroblasts, they observed fewer tumours and decreased inflammation. They also saw reduced expression of the inflammatory molecule IL-6. These differences may be explained in part by how and in which cells the researchers deleted IKK $\beta$ , showing that different subpopulations in the tumour microenvironment have different effects on tumour regulation.



a constitutive gene-deletion approach. Because this leads to much earlier IKK $\beta$  inhibition, the cell population had a chance to adapt to the loss of IKK $\beta$ .

Another potential source of difference lies in the possibility that IKK $\beta$  was deleted in cells other than their targets. Pallangyo and colleagues did not detect gene recombination (indicative of deletion) in epithelial, endothelial or haematopoietic cells, whereas Koliarakis *et al.* have previously reported that their deletion system affects other cell types (chondrocytes, myocytes and keratinocytes)<sup>5</sup> and results in recombination in certain haematopoietic cells. IKK $\beta$  deletion in gut epithelial and myeloid cells can protect against tumour formation<sup>6</sup>, so it is possible that 'unspecific' deletion contributed to the tumour-suppressive effect observed by Koliarakis and colleagues.

Furthermore, different cell markers and experimental tools were used to characterize the targeted cell populations, and the extent of overlap between these cell populations is therefore difficult to assess. It is possible that Pallangyo and colleagues' approach affects most CAFs, whereas Koliarakis and colleagues' method targets just those that are sensitive to

inflammation, which might explain why the latter group sees an effect of IKK $\beta$  deletion only when an inflammatory stimulus is added to the initial mutagen.

Besides these differences, the most exciting aspect of the two studies is that they raise the possibility that fibroblast subpopulations in the tumour stroma may have fundamentally different and even opposing functions in regulating tumour formation and development. Tumour-protective and tumour-promoting characteristics have been attributed to CAFs in various cancers, including pancreatic<sup>1</sup>, skin<sup>7</sup> and mammary cancers<sup>8</sup>. The lack of CAF-specific cell markers and the use of different genetic tools has led to contradictory results and some controversies<sup>1,9</sup>. Although Pallangyo *et al.* and Koliarakis *et al.* both define fibroblasts as collagen-producing cells, the two studies may suffer similar shortcomings.

Future work is needed to better characterize CAFs, and improved genetic tools need to be developed to specifically target distinct stromal cells in mice. If NF- $\kappa$ B–IKK signalling has opposing roles in different fibroblast subpopulations in colorectal cancer, then characterizing these subpopulations, dissecting the underlying mechanisms and extending these

studies to other cancer types will provide new ideas about the role of such cells in cancer. Do these cell populations have potential clinical value? And will CAF-specific proteins ever serve as prognostic markers or targets for anticancer drugs? In my opinion, more basic research is essential to move this exciting field forward, and these two papers are a reminder of the value of multiple parallel studies to build hypotheses on solid experimental data. ■

**Erwin F. Wagner** is at the National Cancer Research Centre (CNIO), Department of Cancer Cell Biology, E-28029 Madrid, Spain. e-mail: ewagner@cnio.es

1. Öhlund, D., Elyada, E. & Tuveson, D. J. *Exp. Med.* **211**, 1503–1523 (2014).
2. Pallangyo, C. K., Ziegler, P. K. & Greten, F. R. *J. Exp. Med.* **212**, 2253–2266 (2015).
3. Koliarakis, V., Pasparakis, M. & Kollias, G. *J. Exp. Med.* **212**, 2235–2251 (2015).
4. DiDonato, J. A., Mercurio, F. & Karin, M. *Immunol. Rev.* **246**, 379–400 (2012).
5. Armaka, M. *et al. J. Exp. Med.* **205**, 331–337 (2008).
6. Greten, F. R. *et al. Cell* **118**, 285–296 (2004).
7. Andreu, P. *et al. Cancer Cell* **17**, 121–134 (2010).
8. Erez, N., Truitt, M., Olson, P., Arron, S. T. & Hanahan, D. *Cancer Cell* **17**, 135–147 (2010).
9. Driskell, R. R. & Watt, F. M. *Trends Cell Biol.* **25**, 92–99 (2015).

## PHYLOGENY

# A home for *Xenoturbella*

Sometimes it is the most unassuming animals that cause the most consternation. *Xenoturbella* (pictured) are simple marine flatworms with no brain, anus, gonads, excretory system or through gut, so one would expect them to find a home among the acoels — similarly simple animals thought to lie at the base of the evolutionary tree of Bilateria, bilaterally symmetrical animals. Yet *Xenoturbella* have caused puzzlement since they were first described in 1949, because quibbles about their ultrastructure and mitochondrial DNA sequences have meant that the worms have never sat entirely happily in their assumed station.

Analysis of nuclear DNA sequences underlined the oddity: *Xenoturbella* were even thought to be highly degenerate molluscs until the revelation that molluscs are what *Xenoturbella* eat. Even stranger was the proposal that *Xenoturbella* and other acoels were most closely related to hemichordates (animals known as acorn worms and pterobranchs) and echinoderms (radially symmetrical marine animals such as sea urchins and starfish). This cast into question the timing of the evolution of several advanced characteristics, such as gill slits, that are shared by members of the deuterostome branch of Bilateria (to which



hemichordates and echinoderms belong), but that are lacking in *Xenoturbella*. It even raised questions about the last common ancestor of Bilateria — perhaps *Xenoturbella* were not as simple as they looked, but had degenerated from a structurally more complex ancestor.

These questions are all but resolved by two studies in this week's issue. Cannon *et al.* (page 89)<sup>1</sup> present a robust phylogenetic analysis based on the gene-transcript profiles of eleven species of *Xenoturbella* and other acoels. This shows that the combined group, known as Xenacoelomorpha, indeed lies

at the very base of the bilaterian radiation. Rouse *et al.* (page 94)<sup>2</sup> add four new species of *Xenoturbella* from the eastern Pacific Ocean to the one already known from the waters of Scotland and Scandinavia. The authors' anatomical and phylogenetic studies on these new forms add weight to the idea that these worms were the earliest to branch from other bilaterians. Zoologists can exhale, and their shy charges can resume their diet of molluscs in peace. **Henry Gee**

1. Cannon, J. T. *et al. Nature* **530**, 89–93 (2016).
2. Rouse, G. W., Wilson, N. G., Carvajal, J. I. & Vrijenhoek, R. C. *Nature* **530**, 94–97 (2016).

# Structural disorder of monomeric $\alpha$ -synuclein persists in mammalian cells

Francois-Xavier Theillet<sup>1†\*</sup>, Andres Binolfi<sup>1†\*</sup>, Beata Bekei<sup>1\*</sup>, Andrea Martorana<sup>2</sup>, Honor May Rose<sup>1</sup>, Marchel Stuver<sup>1</sup>, Silvia Verzini<sup>1</sup>, Dorothea Lorenz<sup>3</sup>, Marleen van Rossum<sup>1</sup>, Daniella Goldfarb<sup>2</sup> & Philipp Selenko<sup>1</sup>

**Intracellular aggregation of the human amyloid protein  $\alpha$ -synuclein is causally linked to Parkinson's disease. While the isolated protein is intrinsically disordered, its native structure in mammalian cells is not known. Here we use nuclear magnetic resonance (NMR) and electron paramagnetic resonance (EPR) spectroscopy to derive atomic-resolution insights into the structure and dynamics of  $\alpha$ -synuclein in different mammalian cell types. We show that the disordered nature of monomeric  $\alpha$ -synuclein is stably preserved in non-neuronal and neuronal cells. Under physiological cell conditions,  $\alpha$ -synuclein is amino-terminally acetylated and adopts conformations that are more compact than when in buffer, with residues of the aggregation-prone non-amyloid- $\beta$  component (NAC) region shielded from exposure to the cytoplasm, which presumably counteracts spontaneous aggregation. These results establish that different types of crowded intracellular environments do not inherently promote  $\alpha$ -synuclein oligomerization and, more generally, that intrinsic structural disorder is sustainable in mammalian cells.**

The effect of the crowded intracellular environment on the structure and dynamics of proteins is poorly understood, which is particularly evident for proteins that lack folded structures in the absence of binding partners, that is, intrinsically disordered proteins (IDPs)<sup>1</sup>. Human  $\alpha$ -synuclein ( $\alpha$ Syn) is a prototypic IDP characterized by its role as the primary protein component of amyloid deposits (termed Lewy bodies) in the brains of patients with Parkinson's disease and other synucleinopathies<sup>2</sup>. While  $\alpha$ Syn is abundantly expressed throughout the brain<sup>3</sup>, amyloid aggregates are primarily found in remnants of apoptotic dopaminergic neurons of the substantia nigra<sup>4</sup>. This raises the possibility that  $\alpha$ Syn adopts different structures in different types of neuronal cells and that these structures exhibit different aggregation propensities<sup>5</sup>. In this light, recent reports postulate that  $\alpha$ Syn principally exists as a folded helical tetramer in intact prokaryotic and eukaryotic cells<sup>6,7</sup>. Although subsequently challenged in several follow-up studies<sup>8–11</sup>, as well as progressively adjusted by the proponents of the initial hypothesis<sup>12–14</sup>, the monomer–tetramer controversy remains central to ongoing discussions about the native structural state(s) of  $\alpha$ Syn in mammalian cells<sup>15</sup>. Addressing this question is crucial for our understanding of possible mechanisms of amyloid formation, not only in the case of  $\alpha$ Syn and Parkinson's disease, but also for the plethora of other neurodegenerative disorders involving IDPs<sup>16</sup>.

## $\alpha$ Syn is disordered in mammalian cells

To obtain atomic-resolution insights into the structure and dynamics of  $\alpha$ Syn in mammalian cells, we performed in-cell NMR experiments in non-neuronal A2780 and HeLa cells, and neuronal B65, SK-N-SH and RCSN-3 cells, the latter being directly derived from rat substantia nigra neurons<sup>17</sup>. To generate in-cell NMR samples in the physiological concentration range of endogenous  $\alpha$ Syn in primary neurons, that is, from 5 to 50  $\mu$ M (refs 3, 18), we developed a simple electroporation

protocol to deliver defined amounts of <sup>15</sup>N isotope-enriched  $\alpha$ Syn into cultured mammalian cells (Fig. 1a). To ensure the physiological intactness of our in-cell NMR samples, we assessed cell viability by flow cytometry and used immunofluorescence and electron microscopy to establish that delivered  $\alpha$ Syn distributed evenly throughout the cytoplasm of electroporated cells (Fig. 1b and Extended Data Fig. 1a, b). In line with previous findings, we measured an intracellular protein half-life of  $\sim$ 50 h (ref. 19) without significant changes in cytoplasmic  $\alpha$ Syn staining (Fig. 1c). To correlate the effective NMR concentrations of isotope-enriched  $\alpha$ Syn in the different cell samples with concentrations that we determined by semi-quantitative western blotting, we recorded one-dimensional (1D) <sup>15</sup>N-filtered in-cell NMR spectra on all specimens (Fig. 2a and Extended Data Fig. 1c). Amide-envelope intensity matching verified that  $90 \pm 10\%$  (mean  $\pm$  s.e.m.) of delivered  $\alpha$ Syn molecules contributed to the measured in-cell NMR signals ( $S_{\text{eff}}$ ), demonstrating that exogenously delivered  $\alpha$ Syn tumbled freely in the cytoplasm of electroporated cells, and did not stably interact with large cellular structures such as membranes. Cell viability and leakage tests after in-cell NMR experiments ruled out signal contributions from external  $\alpha$ Syn (Extended Data Fig. 1d, e).

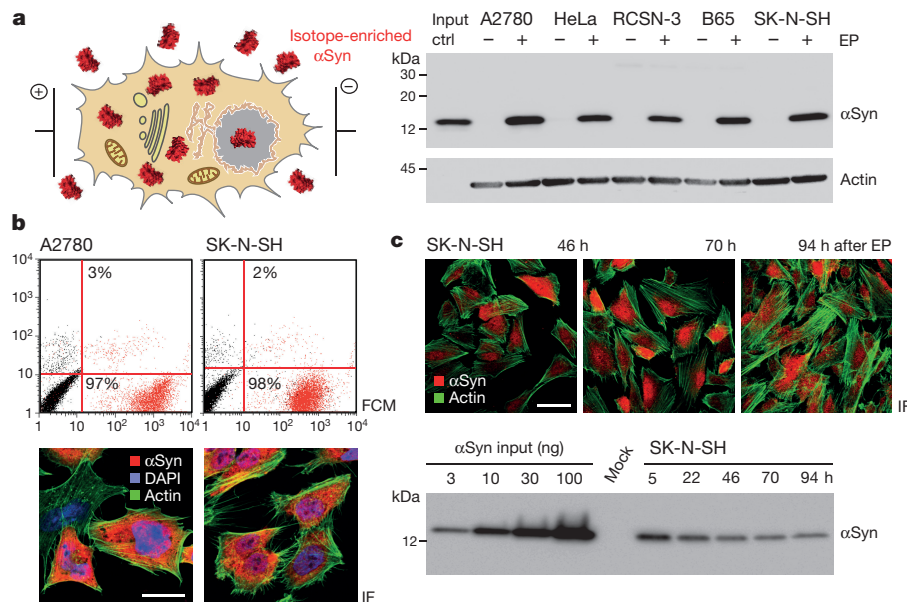
Residue-resolved two-dimensional (2D) <sup>1</sup>H–<sup>15</sup>N in-cell NMR spectra of  $\alpha$ Syn in A2780, HeLa, RCSN-3, B65 and SK-N-SH cells displayed strong similarities with the disordered reference state of the isolated protein, with average backbone amide chemical shift changes ( $\Delta\delta$ ) smaller than 0.01 p.p.m. (Fig. 2b and Extended Data Fig. 1f). These results showed that the different intracellular environments did not induce major conformational rearrangements of monomeric  $\alpha$ Syn. We detected varying degrees of NMR signal broadening in the amino and carboxy terminus of  $\alpha$ Syn, which were highly reproducible in replicate in-cell NMR samples (Extended Data Fig. 2).

<sup>1</sup>In-Cell NMR Laboratory, Department of NMR-supported Structural Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Robert-Rössle Strasse 10, 13125 Berlin, Germany.

<sup>2</sup>Department of Chemical Physics, Weizmann Institute of Science, Rehovot 76100, Israel. <sup>3</sup>Department of Molecular Physiology and Cell Biology, Leibniz Institute of Molecular Pharmacology (FMP Berlin), Robert-Rössle Strasse 10, 13125 Berlin, Germany. <sup>†</sup>Present addresses: Department of Biochemistry, Biophysics and Structural Biology, Institute for Integrative Biology of the Cell (I2BC), UMR9198 CNRS, Bât 144, CEA Saclay, 91191 Gif-sur-Yvette, France (F.-X.T.); Max Planck Laboratory for Structural Biology, Chemistry and Molecular Biophysics of Rosario (MPLBioR-UNR) and Instituto de Investigaciones para el Descubrimiento de Fármacos de Rosario (IIDEFAR-CONICET), 27 de Febrero 210 bis; S2002LRK-Rosario, Argentina (A.B.).

\*These authors contributed equally to this work.





**Figure 1 | Delivery of  $\alpha$ Syn into mammalian cells.** **a**, Electroporation of  $^{15}\text{N}$  isotope-enriched  $\alpha$ Syn into A2780, HeLa, RCSN-3, B65 and SK-N-SH cells delivers 'NMR-visible' protein into 'NMR-invisible' intracellular environments. Western blotting confirms comparable levels of  $\alpha$ Syn transduction. Ctrl; control; EP, electroporated; kDa, kilodaltons. **b**, Top, flow cytometry (FCM) scatter plots of mock-electroporated (black) and Atto488-tagged  $\alpha$ Syn-electroporated (red) A2780 and SK-N-SH cells. Percentages of viable  $\alpha$ Syn-positive (bottom quadrant, x axis) and apoptotic 7-AAD-positive (top quadrant, y axis) cells are indicated. Bottom, immunofluorescence (IF) imaging of  $\alpha$ Syn (red) in electroporated

A2780 and SK-N-SH cells. Phalloidin staining shows actin filaments (green), DAPI staining identifies cell nuclei (blue). Scale bar,  $20\mu\text{m}$ . **c**, Immunofluorescence time course of  $\alpha$ Syn localization and stability in SK-N-SH cells 46, 70 and 94 h after electroporation.  $\alpha$ Syn dilution owing to cell division is compensated with higher laser power settings. Scale bar,  $50\mu\text{m}$ . Semi-quantitative western blotting of intracellular  $\alpha$ Syn at indicated time points after electroporation (total number of electroporated cells loaded). Protein concentrations are determined based on a dilution series of recombinant  $\alpha$ Syn. For gel source data, see Supplementary Fig. 1.

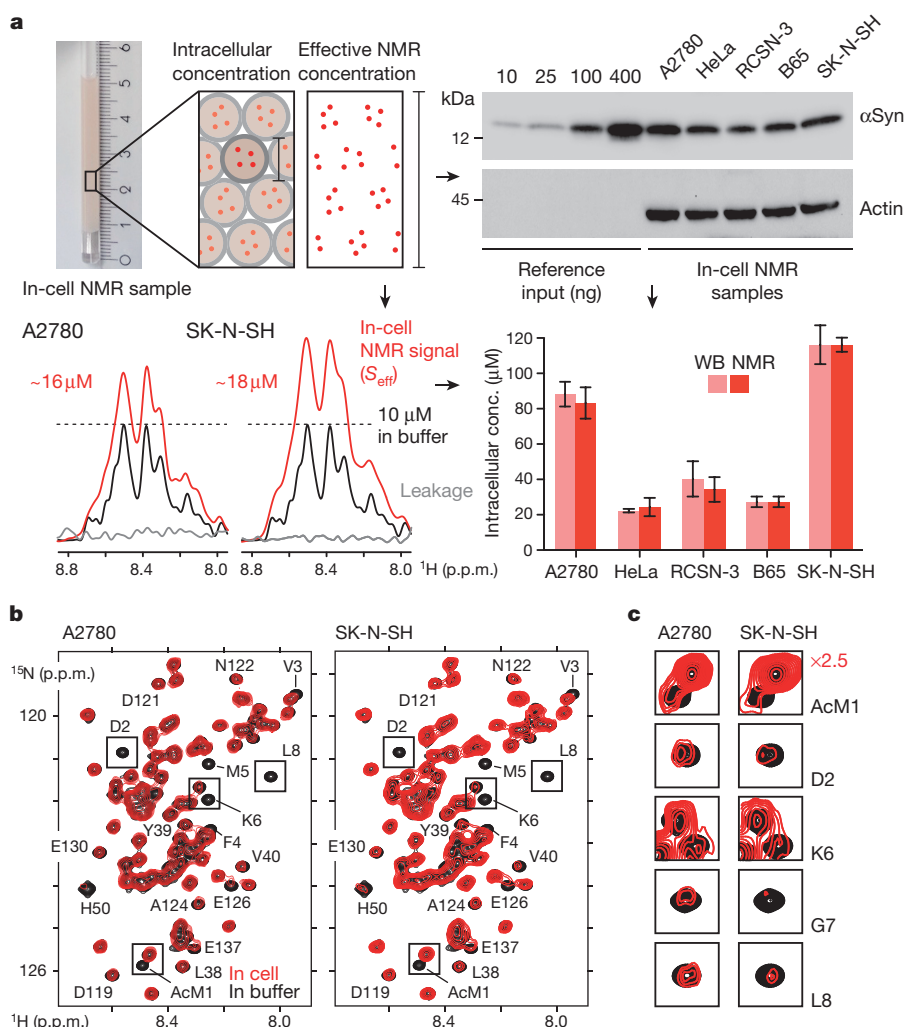
### $\alpha$ Syn is N-terminally acetylated in cells

All in-cell NMR spectra exhibited reduced signal intensities of the first ten residues of  $\alpha$ Syn with peak positions that closely matched those of the N-terminally acetylated protein (Fig. 2c and Extended Data Fig. 3a). Cell lysis obliterated the observed line-broadening effects and clearly revealed the spectral features of N-terminal acetylation (Extended Data Fig. 3b). In further support of the soluble nature of intracellular  $\alpha$ Syn, we quantitatively recovered the monomeric protein in the cytoplasmic fractions of in-cell NMR sample lysates (Extended Data Fig. 3c). Having used electroporation to deliver recombinantly produced, non-acetylated  $\alpha$ Syn into mammalian cells, we reasoned that N-terminal acetylation must have occurred post-translationally. This conclusion challenges the prevalent view that N-terminal acetylation of eukaryotic proteins is established in an exclusive co-translational manner, when nascent polypeptides exit the ribosome<sup>20</sup>. Our findings further confirm that N-terminally acetylated  $\alpha$ Syn represents the physiological form of the protein, in line with previous reports<sup>6,10</sup>. As observed before, acetylation led to higher levels of residual helicity within the N terminus of  $\alpha$ Syn<sup>21,22</sup> and to avoid binding of small unilamellar vesicles (SUVs) that we reconstituted from pig brain polar lipids (Extended Data Fig. 4a, b). Irrespective of N-terminal acetylation, however, in-cell NMR experiments did not reveal the spectral features of fully membrane-associated  $\alpha$ Syn, such as uniform signal broadening of its first  $\sim 100$  residues<sup>22,23</sup>, thus ruling out stable membrane interactions.

### $\alpha$ Syn interacts with the cytoplasm

To characterize the dynamic properties of acetylated  $\alpha$ Syn in the different intracellular environments, we measured in-cell NMR peak intensity changes and backbone amide relaxation parameters in A2780 and SK-N-SH cells (Extended Data Fig. 4c–e), and in artificially crowded solutions containing Ficoll, BSA, lysozyme, SUVs or urea (Extended Data Fig. 5). In cells, signal attenuations primarily affected N- and C-terminal  $\alpha$ Syn residues, whereas a

biologically inert crowding agent such as Ficoll did not recapitulate these line-broadening effects (Fig. 3a). The addition of BSA or lysozyme to N-terminally acetylated  $\alpha$ Syn led to selective reductions of N- or C-terminal signal intensities, respectively, which were reminiscent of the observed in-cell NMR behaviour. Having obtained residue-resolved intracellular  $^{15}\text{N}$  relaxation data (longitudinal ( $R_1$ ), transverse ( $R_2$ ) and  $^1\text{H}$ – $^{15}\text{N}$  hetero-nuclear Overhauser effect (NOE) values), we separated dynamic contributions on the fast nanosecond time scale governing the residue-specific rotational correlation time ( $\tau_c$ ) of  $\alpha$ Syn, from effects in the micro- to millisecond time range giving rise to exchange terms ( $R_{ex}$ ) that also reflect weak transient interactions with cytoplasmic components. Higher  $\tau_c$  profiles reported a uniform decrease of  $\alpha$ Syn dynamics in A2780 and SK-N-SH cells, and in the differently crowded *in vitro* environments as expected for viscosity-driven reductions in overall  $\alpha$ Syn mobility<sup>24</sup> (Fig. 3b and Extended Data Fig. 6a). By contrast, we determined non-uniform exchange contributions in A2780 and SK-N-SH cells, which were largest for the first ten residues of  $\alpha$ Syn and also affected amino acids around Tyr39 and the C terminus of the protein (Fig. 3c and Extended Data Fig. 6b). We detected similar N- or C-terminal exchange profiles with BSA (pI 4.7) or lysozyme (pI 11.35), respectively (Fig. 3c and Extended Data Fig. 6c, d). To test whether complementary electrostatic interactions with the partially charged N and C termini of  $\alpha$ Syn gave rise to these effects, we increased the salt concentrations of BSA- and lysozyme-crowded solutions and re-determined  $R_{ex}$  contributions of  $\alpha$ Syn. We observed no salt effects with BSA, whereas  $R_{ex}$  terms in lysozyme-crowded solutions decreased at higher salt concentrations (Fig. 3c and Extended Data Fig. 6e), suggesting that electrostatic interactions mediated the C-terminal exchange behaviour of  $\alpha$ Syn. Suspecting alternative hydrophobic effects as the cause for exchange contributions in the N terminus of  $\alpha$ Syn, we replaced Phe4 and Tyr39 with alanine residues (that is, F4A;Y39A) and re-measured protein dynamics and  $R_{ex}$  in A2780 and SK-N-SH cells, and in the presence of BSA (Fig. 3c and Extended Data Fig. 7a). We found diminished



**Figure 2 |  $\alpha\text{Syn}$  in-cell NMR samples and spectra.** **a**, In-cell NMR samples and schematic depiction of intracellular protein concentrations versus effective NMR concentrations. Semi-quantitative western blotting (WB) of  $\sim 10^5$  manipulated cells against reference amounts of recombinant  $\alpha\text{Syn}$  delineates average concentrations of delivered protein per cell (calculated in molar terms based on individual cell volumes). 1D  $^1\text{H}$ - $^{15}\text{N}$  NMR spectra of isotope-enriched  $\alpha\text{Syn}$  in A2780 (number of cells in the receiver coil volume ( $N_{\text{cell}}$ ) =  $7 \times 10^7$ ) and SK-N-SH ( $N_{\text{cell}}$  =  $10 \times 10^7$ , 5-mm NMR tubes,  $S_{\text{eff}}$  shown in red) cells, and of 10  $\mu\text{M}$  N-terminally acetylated  $\alpha\text{Syn}$  in buffer (black). 1D traces of sample supernatants (grey) confirm absence of  $\alpha\text{Syn}$  leakage from electroporated cells. Bar graphs depict the comparison of intracellular  $\alpha\text{Syn}$  concentrations determined

by western blotting and NMR (that is,  $[\alpha\text{Syn}]_{\text{cell}}$ ; see Supplementary Methods). Values represent the mean of two independent in-cell NMR samples ( $\alpha\text{Syn}$  concentration in electroporation reactions: 400  $\mu\text{M}$ ). Error bars show the range of highest and lowest values. **b**, 2D  $^1\text{H}$ - $^{15}\text{N}$  NMR spectra of  $\alpha\text{Syn}$  in A2780 and SK-N-SH cells (red, selected region) and of isolated N-terminally acetylated  $\alpha\text{Syn}$  in buffer (black). AcM1, acetylated Met1. **c**, N-terminal  $\alpha\text{Syn}$  residues experiencing site-selective signal attenuations (boxed) are expanded, with in-cell NMR contours plotted at 2.5-fold lower levels (red). In-cell NMR cross-peaks superimpose with reference NMR signals of N-terminally acetylated  $\alpha\text{Syn}$  in buffer (black), confirming the presence of this modification in mammalian cells. For gel source data, see Supplementary Fig. 1.

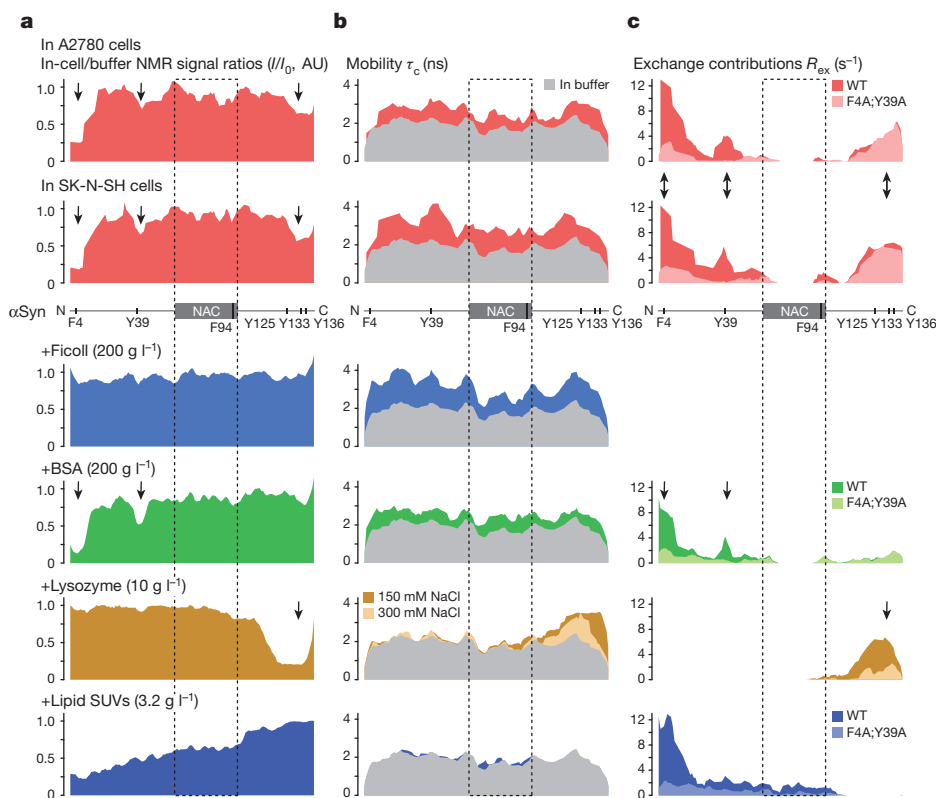
exchange contributions *in vitro* and in cells, which strengthened the notion that hydrophobic contacts mediate N-terminal signal attenuations independently of C-terminal interactions. Binding of N-terminally acetylated  $\alpha\text{Syn}$  (F4A;Y39A) to sub-saturating amounts of SUVs was weaker than that of the wild-type protein, and exchange terms were greatly reduced, although levels of increased residual helicity in the N terminus of mutant  $\alpha\text{Syn}$  were preserved (Fig. 3c and Extended Data Fig. 7b). These results suggested that transient membrane interactions likely contribute to the in-cell NMR behaviour of  $\alpha\text{Syn}$ , especially with regard to hydrophobic ‘membrane scanning’ of N-terminal  $\alpha\text{Syn}$  residues<sup>25</sup>. Because cell lysis obliterated the observed line broadening effects (Extended Data Fig. 3b), we concluded that weak hydrophobic and electrostatic interactions are not preserved under dilute extract conditions. Thus, we demonstrate that protein dynamics of monomeric  $\alpha\text{Syn}$  in intact mammalian cells are governed by viscosity-driven reductions in overall mobility and by weak hydrophobic and electrostatic interactions of N- and C-terminal residues

with intracellular components, respectively (Fig. 4a). Notably, our data also revealed that the aggregation-prone non-amyloid- $\beta$  component (NAC) region of  $\alpha\text{Syn}$  experiences the least effects in the different intracellular environments, without substantial exchange contributions (Fig. 3c). While this may reflect the paucity of aromatic and charged residues in the NAC region, it may also indicate the presence of intracellular protein conformations that shield NAC residues from exposure to the cytoplasm (Fig. 4b).

### $\alpha\text{Syn}$ adopts compact structures in cells

To investigate whether such  $\alpha\text{Syn}$  conformations exist in cells, we performed NMR paramagnetic relaxation enhancement (PRE) and electron paramagnetic resonance (EPR), double electron-electron resonance (DEER) measurements<sup>26</sup> on A2780 and SK-N-SH cells that we electroporated with appropriately isotope-enriched and stably maleimide-DOTA-tagged and Gd(III)-loaded  $\alpha\text{Syn}$  (Fig. 5a and Extended Data Figs 8 and 9a, b). Previous PRE experiments on isolated  $\alpha\text{Syn}$  showed





**Figure 3 |  $\alpha$ Syn dynamics in cells and crowded solutions.** **a**, Residue-resolved NMR signal intensity ratios ( $I/I_0$ ) of  $\alpha$ Syn in A2780 and SK-N-SH cells (red) identify regions of site-selective line broadening (marked by arrows). No line broadening of N-terminally acetylated  $\alpha$ Syn is observed in the presence of Ficoll (blue), whereas BSA (green) and lysozyme (orange) recapitulate signal attenuations of N- and C-terminal residues, respectively. Addition of sub-saturating amounts of SUVs to N-terminally acetylated  $\alpha$ Syn leads to a gradual reduction of NMR signal intensities of its first 100 residues. For simplification, all profiles show values averaged over three consecutively resolved residues. AU, arbitrary units. **b**, Residue-resolved rotational correlation time ( $\tau_c$ ) of N-terminally acetylated  $\alpha$ Syn in buffer (grey), in cells (red), and in the presence of different crowding agents (averaged over three consecutively resolved residues) reveal

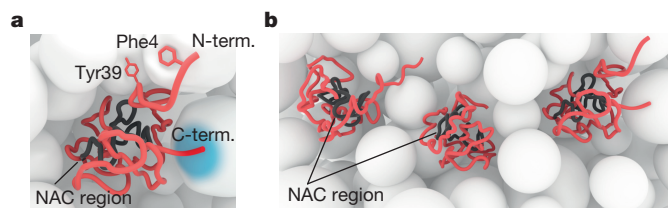
that the protein exhibits transient long-range interactions between N- and C-terminal residues<sup>27,28</sup>. These intramolecular contacts result in loosely packed  $\alpha$ Syn structures that deviate from extended polypeptide chain conformations and occlude the central NAC region. PRE experiments on Gd(III)-labelled  $\alpha$ Syn in A2780 and SK-N-SH cells revealed distance profiles that were similar to those of the isolated protein, thus arguing for the preservation of intramolecular  $\alpha$ Syn contacts in intact

uniform reductions of fast protein motions. **c**, Site-specific exchange contributions ( $R_{ex}$ ) identify  $\alpha$ Syn regions that engage in weak transient interactions with cytoplasmic components (marked with arrows). Substitution of Phe4 and Tyr39 with alanine residues (F4A;Y39A) reduces N-terminal  $R_{ex}$  terms in cells (light red) and in BSA-crowded solutions (light green), thus identifying hydrophobic contacts as the source of N-terminal exchange contributions. Higher salt concentrations diminish C-terminal  $\alpha$ Syn interactions with lysozyme (light orange), arguing for electrostatic effects. Different from BSA-crowded solutions, only residues 1–20 of N-terminally acetylated  $\alpha$ Syn exhibit pronounced exchange contributions in the presence of sub-saturating amounts of SUVs (blue), which are diminished for  $\alpha$ Syn(F4A;Y39A) (light blue).

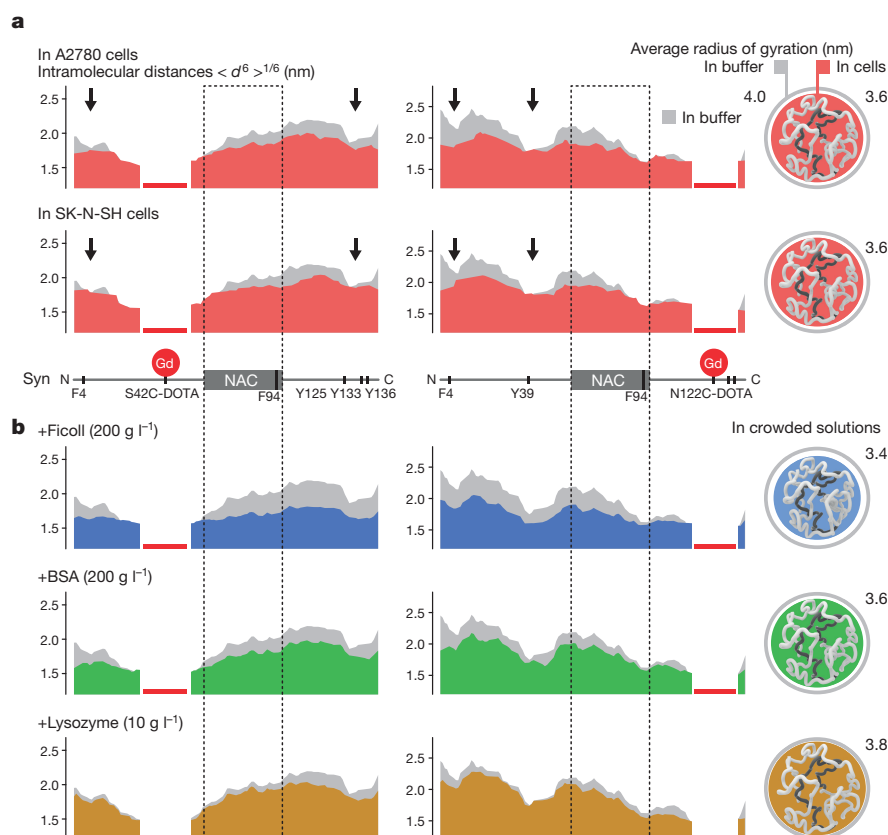
cells (Fig. 5a and Extended Data Fig. 8). Notably, however, in-cell PRE profiles also indicated greater levels of  $\alpha$ Syn compaction, which suggested that both intracellular environments strengthen, rather than weaken, the structural features that  $\alpha$ Syn displays in buffer. PRE experiments in Ficoll- and BSA-crowded solutions (200 g l<sup>-1</sup>) recapitulated these compaction effects (Fig. 5b), whereas denaturation with 8 M urea led to extended  $\alpha$ Syn structures (Extended Data Fig. 8). Complementary DEER experiments in A2780 cells confirmed these results (Extended Data Fig. 9c, d). To test whether  $\alpha$ Syn formed oligomers or aggregates in A2780 cells, we also performed intermolecular DEER and PRE measurements<sup>29</sup>. Even at intracellular  $\alpha$ Syn concentrations above 50  $\mu$ M, neither experimental approach yielded indications for such events, although we detected sparsely populated  $\alpha$ Syn oligomers in the presence of 200 g l<sup>-1</sup> Ficoll and BSA (Extended Data Fig. 9e, f).

## Discussion

Our results show that exogenously delivered  $\alpha$ Syn exists as an N-terminally acetylated, disordered and highly dynamic monomer in neuronal and non-neuronal cells, without detectable signs of oligomerization, spontaneous aggregation, or targeted degradation. In the absence of chemical cross-linking<sup>12–14</sup> and the concomitant introduction of oligomer-promoting agents such as dimethyl sulfoxide (DMSO)<sup>30</sup>, monomeric  $\alpha$ Syn represents the predominant species in the cytoplasm of cells analysed in this study. Although we cannot



**Figure 4 |  $\alpha$ Syn interactions and conformations in cells.** **a**, Cartoon representation of localized hydrophobic and electrostatic interactions of  $\alpha$ Syn with cytoplasmic components, shown schematically as white spheres. Hydrophobic contacts are mediated by aromatic residues Phe4 and Tyr39 in the N terminus of  $\alpha$ Syn whereas negatively charged residues in the C terminus engage in complementary electrostatic interactions with positively charged intracellular surfaces (indicated in blue). The central NAC region (dark grey) is shielded from exposure to the cytoplasm and does not interact with intracellular components. **b**, Cartoon models of possible ensemble conformations of monomeric  $\alpha$ Syn in cells.



**Figure 5 | Compact  $\alpha$ Syn structures in cells and crowded solutions.** **a**, Intramolecular PRE-derived distance profiles of N-terminally acetylated  $\alpha$ Syn in buffer (grey) and in A2780 and SK-N-SH cells (red). **b**, Comparative PRE-derived distance profiles in Ficoll- (blue), BSA- (green) and lysozyme- (orange) crowded solutions. Gd(III)-DOTA tags at residues 42 (left; S42C-DOTA) and 122 (right; N122C-DOTA) provide complementary information about  $\alpha$ Syn compaction in the respective environments. Proximal PRE effects render residues adjacent to the conjugated Gd(III) invisible (red bars). PRE-derived distances were obtained assuming that every Gd(III)-<sup>1</sup>H vector fluctuation rate scales

linearly with that of the backbone N-H vector and the Gd(III)-DOTA complex. For simplification, profiles show values averaged over three consecutively resolved residues. Comparisons of  $\alpha$ Syn dimensions in buffer, in cells and in the presence of different crowding agents are shown on the right. Average radii of gyration and levels of  $\alpha$ Syn compaction are delineated based on the scaling of representative PRE distances relative to values measured in buffer (see Supplementary Methods and Extended Data Fig. 8). In the depicted cartoon models, residues of the NAC region are coloured in dark grey. Arrows denote regions of marked intramolecular contacts.

rule out the presence of other lowly populated  $\alpha$ Syn states in our in-cell NMR and EPR samples, we can exclude scenarios in which most  $\alpha$ Syn molecules adopt stably folded, or fully membrane-associated structures<sup>6,9</sup>. In-cell conformations of  $\alpha$ Syn are generally more compact than in buffer, and shield hydrophobic residues of the amyloidogenic NAC region from interactions with the cytoplasm, similar to the folding principle of structured proteins. By contrast, Phe4 and Tyr39 prominently engage in transient interactions with cytoplasmic components, whereas negatively charged residues in the  $\alpha$ Syn C terminus participate in weak electrostatic contacts. Both types of interactions can be recapitulated in artificially crowded *in vitro* solutions and are lost after cell lysis, reminiscent of previously observed quinary structure interactions of folded and partially disordered proteins in prokaryotic and eukaryotic cells<sup>31,32</sup>. By exhibiting structural features that disfavour NAC-mediated interactions, large conformational rearrangements appear to be necessary for  $\alpha$ Syn to oligomerize under native cell conditions<sup>33</sup>, for which the identified sampling of hydrophobic contacts by Phe4 and Tyr39, including the transient binding of cellular membranes may set the stage<sup>34</sup>. Given that these residues are critical for  $\alpha$ Syn aggregation *in vitro*<sup>35</sup>, they may also constitute ‘interaction hotspots’ in the formation of early  $\alpha$ Syn oligomers in cells.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 June; accepted 14 December 2015.

Published online 25 January 2016.

1. Theillet, F. X. *et al.* Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chem. Rev.* **114**, 6661–6714 (2014).
2. Goedert, M., Spillantini, M. G., Del Tredici, K. & Braak, H. 100 years of Lewy pathology. *Nature Rev. Neurol.* **9**, 13–24 (2013).
3. Iwai, A. *et al.* The precursor protein of non-A beta component of Alzheimer's disease amyloid is a presynaptic protein of the central nervous system. *Neuron* **14**, 467–475 (1995).
4. Lashuel, H. A., Overk, C. R., Oueslati, A. & Masliah, E. The many faces of alpha-synuclein: from structure and toxicity to therapeutic target. *Nature Rev. Neurosci.* **14**, 38–48 (2013).
5. Dettmer, U., Selkoe, D. & Bartels, T. New insights into cellular  $\alpha$ -synuclein homeostasis in health and disease. *Curr. Opin. Neurobiol.* **36**, 15–22 (2015).
6. Bartels, T., Choi, J. G. & Selkoe, D. J.  $\alpha$ -Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature* **477**, 107–110 (2011).
7. Wang, W. *et al.* A soluble  $\alpha$ -synuclein construct forms a dynamic tetramer. *Proc. Natl Acad. Sci. USA* **108**, 17797–17802 (2011).
8. Binolfi, A., Theillet, F. X. & Selenko, P. Bacterial in-cell NMR of human alpha-synuclein: a disordered monomer by nature? *Biochem. Soc. Trans.* **40**, 950–954 (2012).
9. Burré, J. *et al.* Properties of native brain  $\alpha$ -synuclein. *Nature* **498**, E4–E6 (2013).
10. Fauvet, B. *et al.* Characterization of semisynthetic and naturally N-alpha-acetylated  $\alpha$ -synuclein *in vitro* and in intact cells: implications for aggregation and cellular properties of alpha-synuclein. *J. Biol. Chem.* **287**, 28243–28262 (2012).
11. Fauvet, B. *et al.*  $\alpha$ -Synuclein in central nervous system and from erythrocytes, mammalian cells, and *Escherichia coli* exists predominantly as disordered monomer. *J. Biol. Chem.* **287**, 15345–15364 (2012).



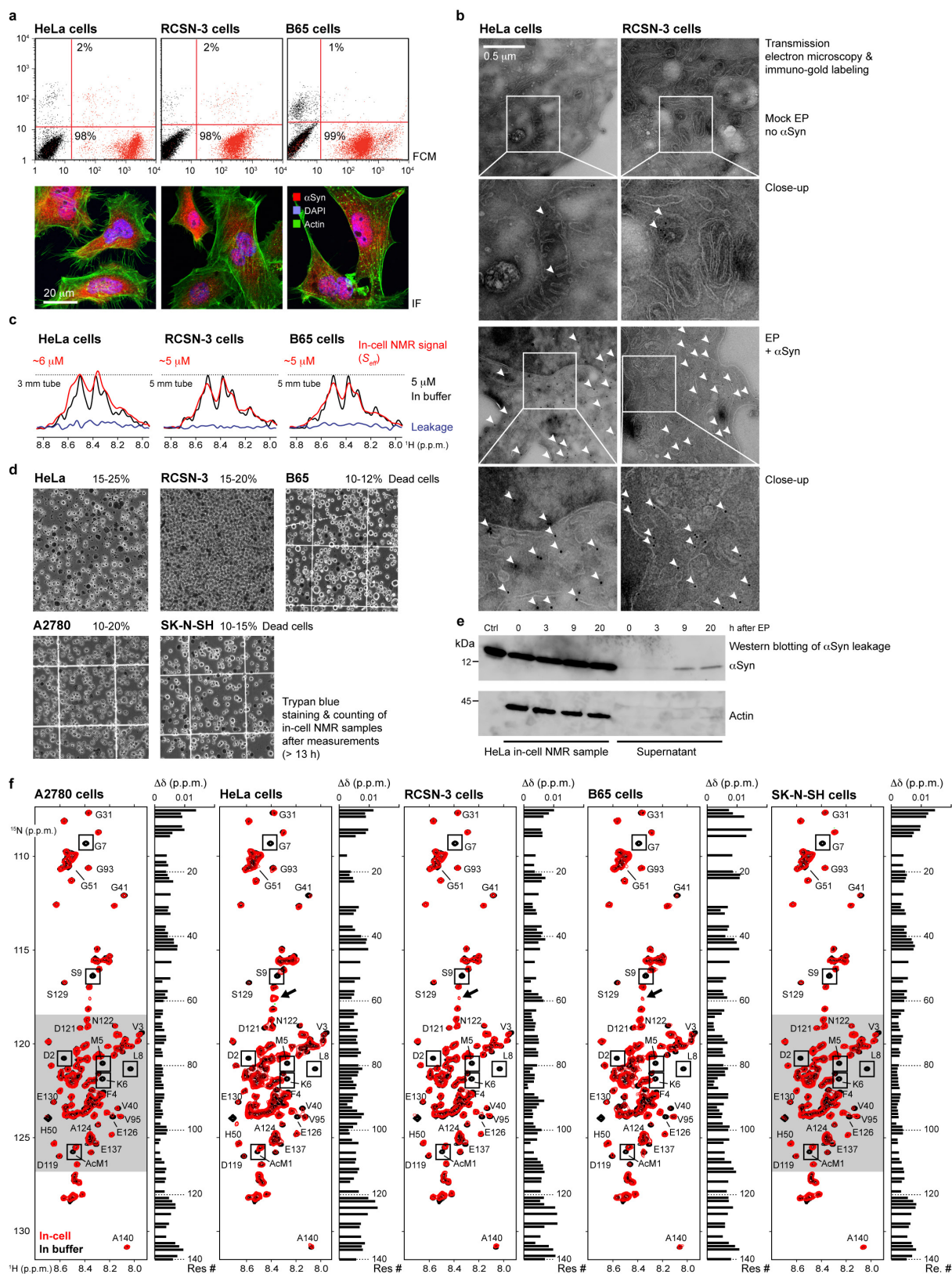
12. Dettmer, U., Newman, A. J., Luth, E. S., Bartels, T. & Selkoe, D. *In vivo* cross-linking reveals principally oligomeric forms of  $\alpha$ -synuclein and  $\beta$ -synuclein in neurons and non-neural cells. *J. Biol. Chem.* **288**, 6371–6385 (2013).
13. Dettmer, U. *et al.* Parkinson-causing alpha-synuclein missense mutations shift native tetramers to monomers as a mechanism for disease initiation. *Nature Commun.* **6**, 7314 (2015).
14. Luth, E. S., Bartels, T., Dettmer, U., Kim, N. C. & Selkoe, D. J. Purification of  $\alpha$ -synuclein from human brain reveals an instability of endogenous multimers as the protein approaches purity. *Biochemistry* **54**, 279–292 (2015).
15. Selkoe, D. *et al.* Defining the native state of  $\alpha$ -synuclein. *Neurodegener. Dis.* **13**, 114–117 (2014).
16. Uversky, V. N. Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Front. Aging Neurosci.* **7**, 18 (2015).
17. Paris, I. *et al.* The catecholaminergic RCSN-3 cell line: a model to study dopamine metabolism. *Neurotox. Res.* **13**, 221–230 (2008).
18. Wilhelm, B. G. *et al.* Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. *Science* **344**, 1023–1028 (2014).
19. Okochi, M. *et al.* Constitutive phosphorylation of the Parkinson's disease associated  $\alpha$ -synuclein. *J. Biol. Chem.* **275**, 390–397 (2000).
20. Starheim, K. K., Gevaert, K. & Arnesen, T. Protein N-terminal acetyltransferases: when the start matters. *Trends Biochem. Sci.* **37**, 152–161 (2012).
21. Kang, L. *et al.* N-terminal acetylation of alpha-synuclein induces increased transient helical propensity and decreased aggregation rates in the intrinsically disordered monomer. *Protein Sci.* **21**, 911–917 (2012).
22. Maltsev, A. S., Ying, J. F. & Bax, A. Impact of N-terminal acetylation of  $\alpha$ -synuclein on its random coil and lipid binding properties. *Biochemistry* **51**, 5004–5013 (2012).
23. Dikiy, I. & Eliezer, D. N-terminal acetylation stabilizes N-terminal helicity in lipid- and micelle-bound alpha-synuclein and increases its affinity for physiological membranes. *J. Biol. Chem.* **289**, 3652–3665 (2014).
24. Li, C. *et al.* Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: Implications for in-cell NMR spectroscopy. *J. Am. Chem. Soc.* **130**, 6310–6311 (2008).
25. Fusco, G. *et al.* Direct observation of the three regions in  $\alpha$ -synuclein that determine its membrane-bound behaviour. *Nature Commun.* **5**, 3827 (2014).
26. Martorana, A. *et al.* Probing protein conformation in cells by EPR distance measurements using Gd<sup>3+</sup> spin labelling. *J. Am. Chem. Soc.* **136**, 13458–13465 (2014).
27. Bertocini, C. W. *et al.* Release of long-range tertiary interactions potentiates aggregation of natively unstructured  $\alpha$ -synuclein. *Proc. Natl Acad. Sci. USA* **102**, 1430–1435 (2005).
28. Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. Mapping long-range interactions in  $\alpha$ -synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J. Am. Chem. Soc.* **127**, 476–477 (2005).
29. Wu, K. P. & Baum, J. Detection of transient interchain interactions in the intrinsically disordered protein  $\alpha$ -synuclein by NMR paramagnetic relaxation enhancement. *J. Am. Chem. Soc.* **132**, 5546–5547 (2010).
30. Kostka, M. *et al.* Single particle characterization of iron-induced pore-forming  $\alpha$ -synuclein oligomers. *J. Biol. Chem.* **283**, 10992–11003 (2008).
31. Danielsson, J. *et al.* Thermodynamics of protein destabilization in live cells. *Proc. Natl Acad. Sci. USA* **112**, 12402–12407 (2015).
32. Monteith, W. B., Cohen, R. D., Smith, A. E., Guzman-Cisneros, E. & Pielak, G. J. Quinary structure modulates protein stability in cells. *Proc. Natl Acad. Sci. USA* **112**, 1739–1742 (2015).
33. Cremades, N. *et al.* Direct observation of the interconversion of normal and toxic forms of  $\alpha$ -synuclein. *Cell* **149**, 1048–1059 (2012).
34. Galvagnion, C. *et al.* Lipid vesicles trigger  $\alpha$ -synuclein aggregation by stimulating primary nucleation. *Nature Chem. Biol.* **11**, 229–234 (2015).
35. Lamberto, G. R. *et al.* Structural and mechanistic basis behind the inhibitory interaction of PcTS on  $\alpha$ -synuclein amyloid fibril formation. *Proc. Natl Acad. Sci. USA* **106**, 21057–21062 (2009).
36. Ulmer, T. S. & Bax, A. Comparison of structure and dynamics of micellebound human  $\alpha$ -synuclein and Parkinson disease variants. *J. Biol. Chem.* **280**, 43179–43187 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. Burre, T. Sudhof, D. Jovin, D. Mulvihill, P. Caviedes and T. Maritzen for plasmids and reagents, and for critical discussions. V. Subramanian, D. Eliezer, J. Clark and J. Dyson for carefully reading the manuscript and helpful comments. M. Herzig for essential input in the development of the electroporation protocol and M. Beerbaum and P. Schmieder for maintenance of the NMR infrastructure. RCSN-3 cells are made available by P. Caviedes upon request (pcaviede@med.uchile.cl). F.-X.T. acknowledges funding from the Association pour la Recherche sur le Cancer (ARC). D.G. acknowledges support by the Israel Science Foundation (ISF) F.I.R.S.T. program (grant 1114/12). P.S. was funded by the Deutsche Forschungsgemeinschaft (DFG) via an Emmy Noether Project Grant (SE1794/1-1). This work is supported by the European Research Council (ERC) Consolidator Grant (CoG) 'NeuroInCellNMR' (647474) awarded to P.S.

**Author Contributions** F.-X.T. and A.B. performed in-cell and *in vitro* NMR experiments. B.B., H.M.R., S.V., M.v.R. and M.S. developed the electroporation protocol, manipulated and handled cells, and performed light microscopy, cell and molecular biology experiments. A.M. and D.G. designed and performed EPR experiments, and analysed data. D.L. performed electron microscopy experiments. F.-X.T., A.B. and P.S. conceived the study, analysed the data and wrote the paper. F.-X.T., A.B. and B.B. contributed equally. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.S. ([selenko@fmp-berlin.de](mailto:selenko@fmp-berlin.de)).



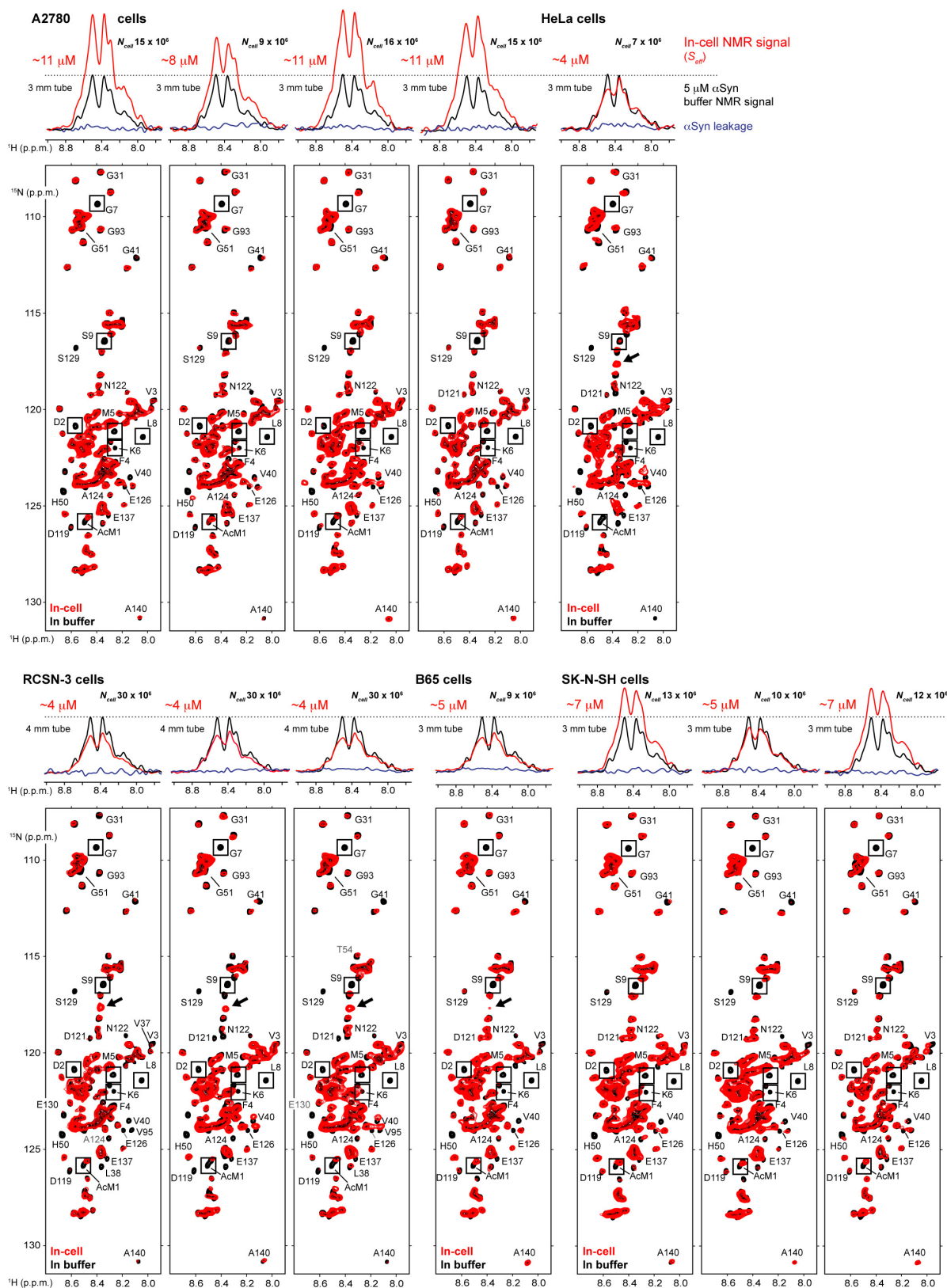
Extended Data Figure 1 | See next page for figure caption.



**Extended Data Figure 1 | Quality control experiments and full 2D in-cell NMR spectra.**

**a**, Flow cytometry scatter-plots of electroporated HeLa, RCSN-3 and B65 cells carrying Atto488-tagged  $\alpha$ Syn (red,  $x$  axis) counterstained with 7-AAD ( $y$  axis). Percentages of viable  $\alpha$ Syn-positive (bottom quadrant) and apoptotic (top quadrant) cells are indicated. Mock-electroporated cells are shown in black. Immunofluorescence imaging of  $\alpha$ Syn (red) in electroporated HeLa, RCSN-3 and B65 cells. Phalloidin staining shows actin filaments (green), DAPI staining identifies cell nuclei (blue). **b**, Cryo-electron microscopy and anti- $\alpha$ Syn immunogold-labelling of mock-electroporated (top) and  $\alpha$ Syn-electroporated (bottom) HeLa and RCSN-3 cells confirms organelle intactness and uniform distribution of delivered  $\alpha$ Syn. Arrowheads indicate positions of gold-particles indicative of intracellular  $\alpha$ Syn. **c**, Representative 1D  $^1\text{H}$ - $^{15}\text{N}$  amide-envelope traces of  $\alpha$ Syn in-cell NMR signals ( $S_{\text{eff}}$  shown in red) in HeLa ( $N_{\text{cell}} = 2 \times 10^7$ ), RCSN-3 ( $N_{\text{cell}} = 7 \times 10^7$ ) and B65 ( $N_{\text{cell}} = 7 \times 10^7$ ) cells (electroporated with  $400 \mu\text{M}$   $\alpha$ Syn). Reference traces of  $5 \mu\text{M}$   $^{15}\text{N}$  isotope-enriched, N-terminally acetylated  $\alpha$ Syn in buffer are shown in black. 1D NMR traces of leakage tests after the respective in-cell NMR experiments, acquired with the

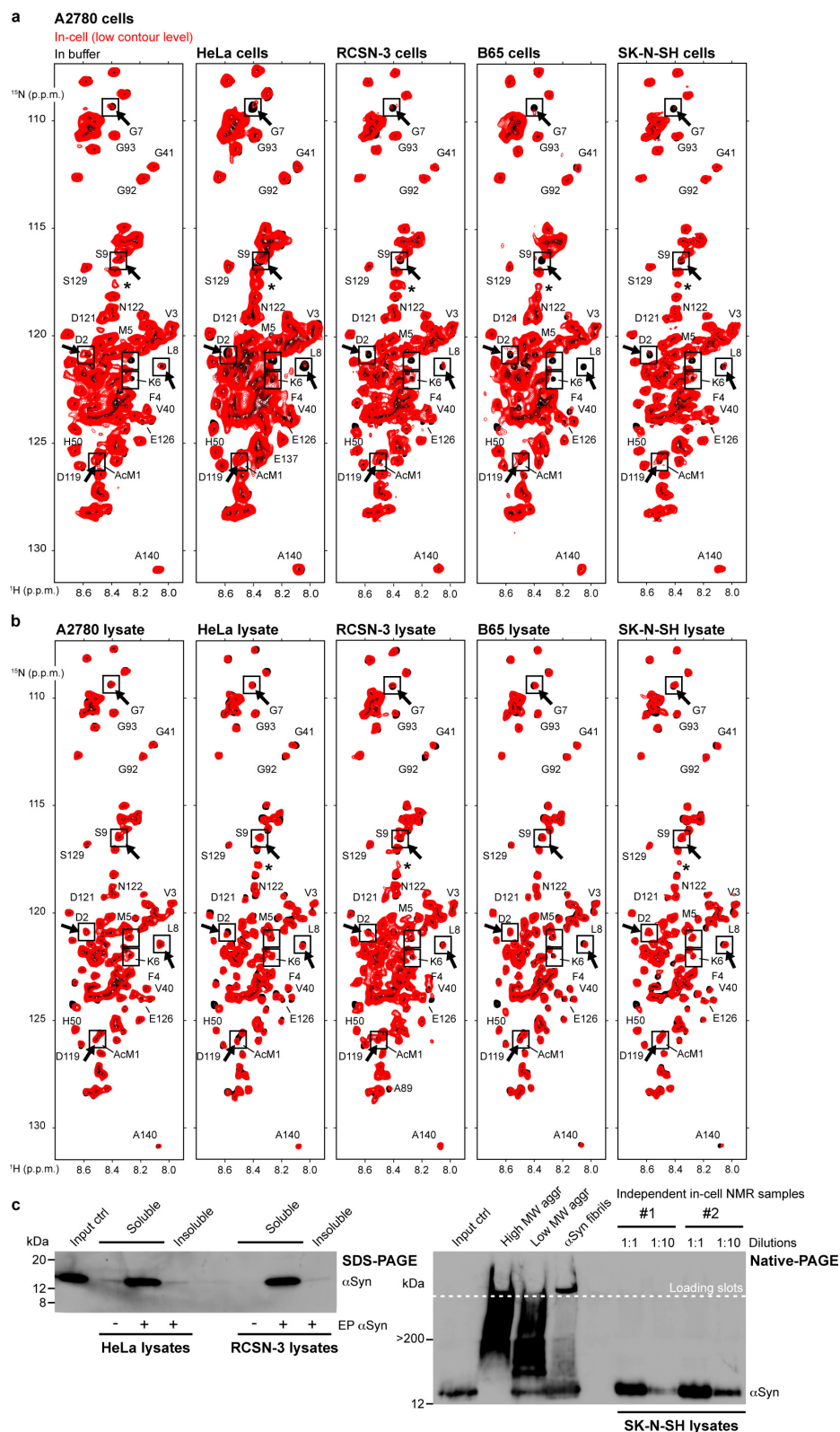
same NMR acquisition parameters and spectrometer settings as for the in-cell NMR samples, are shown in blue (see Supplementary Methods). **d**, Exemplary Trypan blue cell-viability tests on in-cell NMR specimens. Average percentages of dead cells are indicated for the different cell lines. **e**, Time course western blot of  $\alpha$ Syn leakage from HeLa cells (that is, highest percentage of dead cells after in-cell NMR experiments) reveals negligible amounts of extracellular protein. **f**, Overlay of 2D in-cell NMR spectra of  $\alpha$ Syn in A2780, HeLa, RCSN-3, B65 and SK-N-SH cells (red, full spectral region) and of N-terminally acetylated  $\alpha$ Syn in buffer (black). Selected regions of NMR spectra shown in Fig. 2b are shaded in grey. Sites of N-terminal line broadening are boxed. Arrows denote growth medium-specific metabolite signals (background). Chemical shift differences of non-proline backbone amide ( $^1\text{H}$ - $^{15}\text{N}$ ) resonances between  $\alpha$ Syn in buffer and in cells were calculated as  $\Delta\delta = [(\Delta\delta_{^1\text{H}})^2 + (\Delta\delta_{^{15}\text{N}} \times 0.2)^2]^{1/2}$  (in p.p.m.) and are shown next to the corresponding in-cell NMR spectra. Residues without  $\Delta\delta$  values were not analysed owing to spectral overlap.  $\Delta\delta$  of His50 is not included because of the known pH sensitivity of this resonance signal. For gel source data, see Supplementary Fig. 1.



**Extended Data Figure 2 | Reproducibility of in-cell NMR samples and spectra.** Representative in-cell NMR spectra (red) of independent replicate samples in the different mammalian cell types (electroporation concentrations of  $\alpha\text{Syn}$ : 400  $\mu\text{M}$ ). 1D  $^1\text{H}$ - $^{15}\text{N}$  traces of in-cell amide-envelope signals ( $S_{\text{eff}}$  red) are shown above the respective 2D NMR

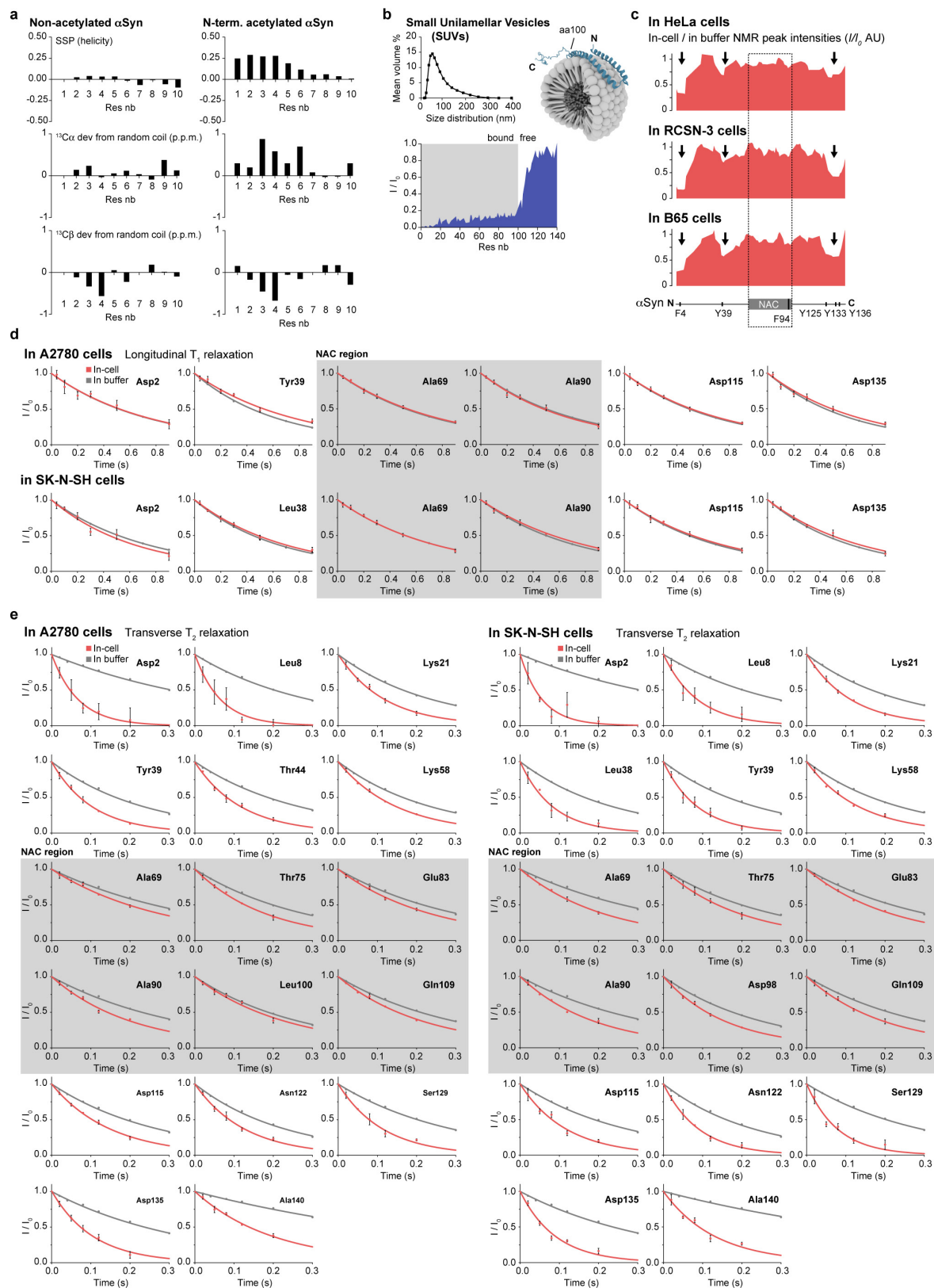
spectra (5  $\mu\text{M}$  reference trace: black; leakage trace: blue).  $N_{\text{cell}}$  and NMR tube diameters are indicated. 2D reference NMR spectra of N-terminally acetylated  $\alpha\text{Syn}$  (5  $\mu\text{M}$ ) are shown in black. Sites of N-terminal line broadening are boxed. Growth medium-specific natural abundance metabolite signals (background) are marked with arrows.





**Extended Data Figure 3 | Low contour level in-cell and lysate NMR spectra.** **a**, Same in-cell NMR spectra as in Extended Data Fig. 1f plotted at lower contour levels (red). Boxes and arrows indicate broadened NMR signals of N-terminal αSyn residues, including acetylated Met1. N-terminally acetylated αSyn (5 μM) is shown in the respective reference NMR spectra (black). Asterisks denote natural abundance signals of background metabolites (growth medium specific). **b**, Same samples as in **a** after cell lysis. NMR spectra were directly acquired on the soluble fractions of cleared lysates (red), alleviating N-terminal line broadening and displaying signal overlap with the acetylated reference state of αSyn.

Chemical shift changes of His50 indicate minor pH changes. **c**, Left, fractionation, SDS-PAGE separation and western blotting of HeLa and RCSN-3 in-cell NMR sample lysates indicate recovery of intracellular αSyn in soluble lysate fractions. Right, separation of SK-N-SH αSyn in-cell NMR sample lysates by native PAGE yields monomeric αSyn. Covalent low- and high-molecular mass (MW) cytochrome *c*/H<sub>2</sub>O<sub>2</sub> aggregates of αSyn and non-covalent αSyn fibrils serve as input controls for differently aggregated forms of the protein (see Supplementary Methods). For gel source data, see Supplementary Fig. 1.

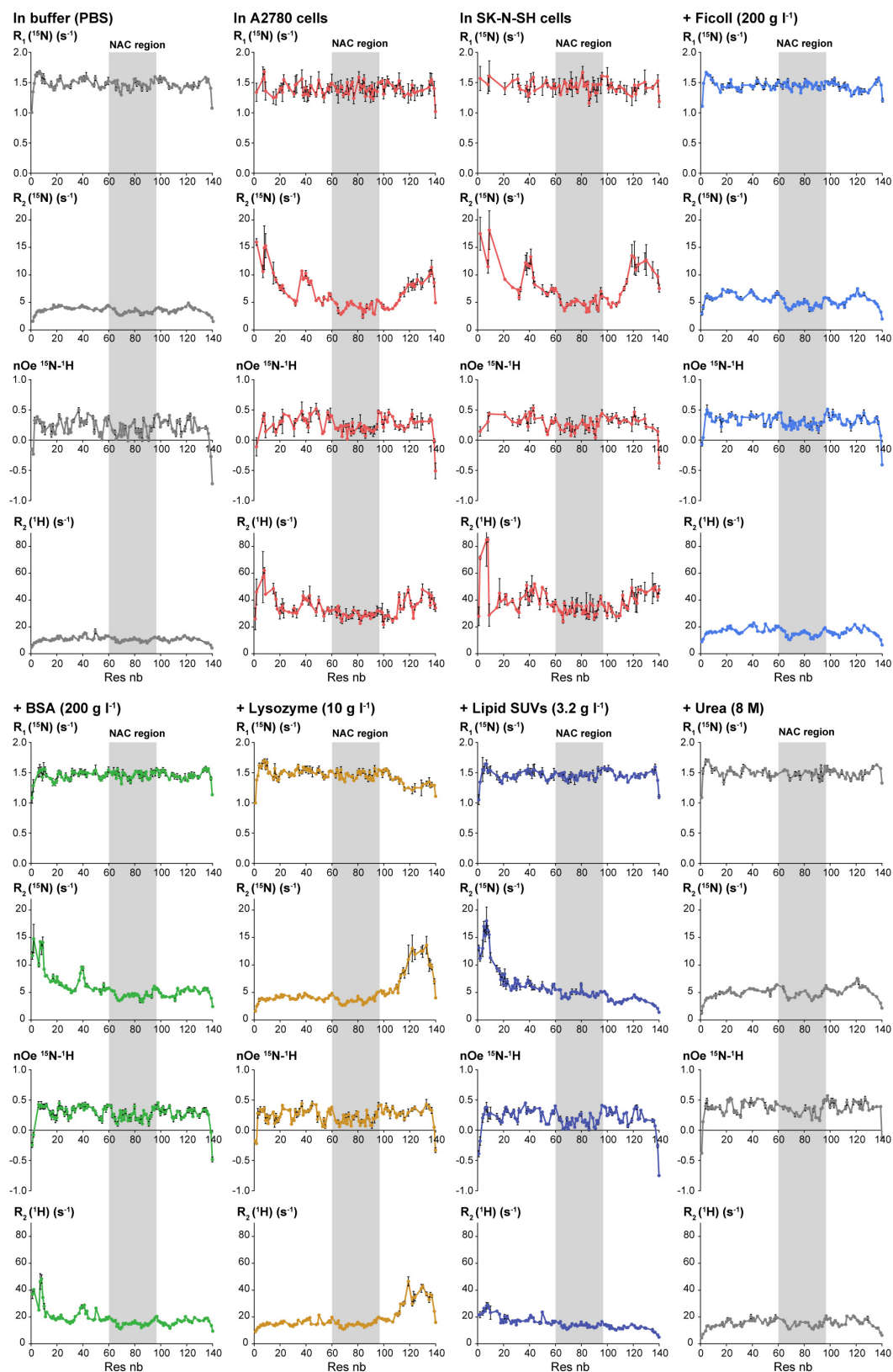


Extended Data Figure 4 | See next page for figure caption.



**Extended Data Figure 4 | NMR characterization of N-terminally acetylated  $\alpha$ Syn and in-cell relaxation data.** **a**, Secondary structure propensity (SSP) scores and  $C\alpha$ ,  $C\beta$  chemical shift deviations from random coil values indicate levels of residual helicity within the first ten residues of non-acetylated (left) and N-terminally acetylated (right)  $\alpha$ Syn. **b**, Dynamic light scattering (DLS) of SUVs prepared from pig brain polar lipids (average diameter  $\sim 60$  nm) and schematic depiction of one possible  $\alpha$ Syn conformation when bound to SUVs (according to ref. 36).  $\alpha$ Syn membrane binding is mediated by its first  $\sim 100$  residues forming either a continuous or a broken  $\alpha$ -helix. Residue-resolved signal intensity ratios ( $I/I_0$ ) of free  $\alpha$ Syn ( $I_0$ ) versus  $\alpha$ Syn in the presence of SUVs ( $I$ ), obtained with  $^{15}\text{N}$  isotope-enriched, N-terminally acetylated  $\alpha$ Syn and saturating amounts of SUVs ( $16\text{ g l}^{-1}$ ).  $\alpha$ Syn membrane binding results

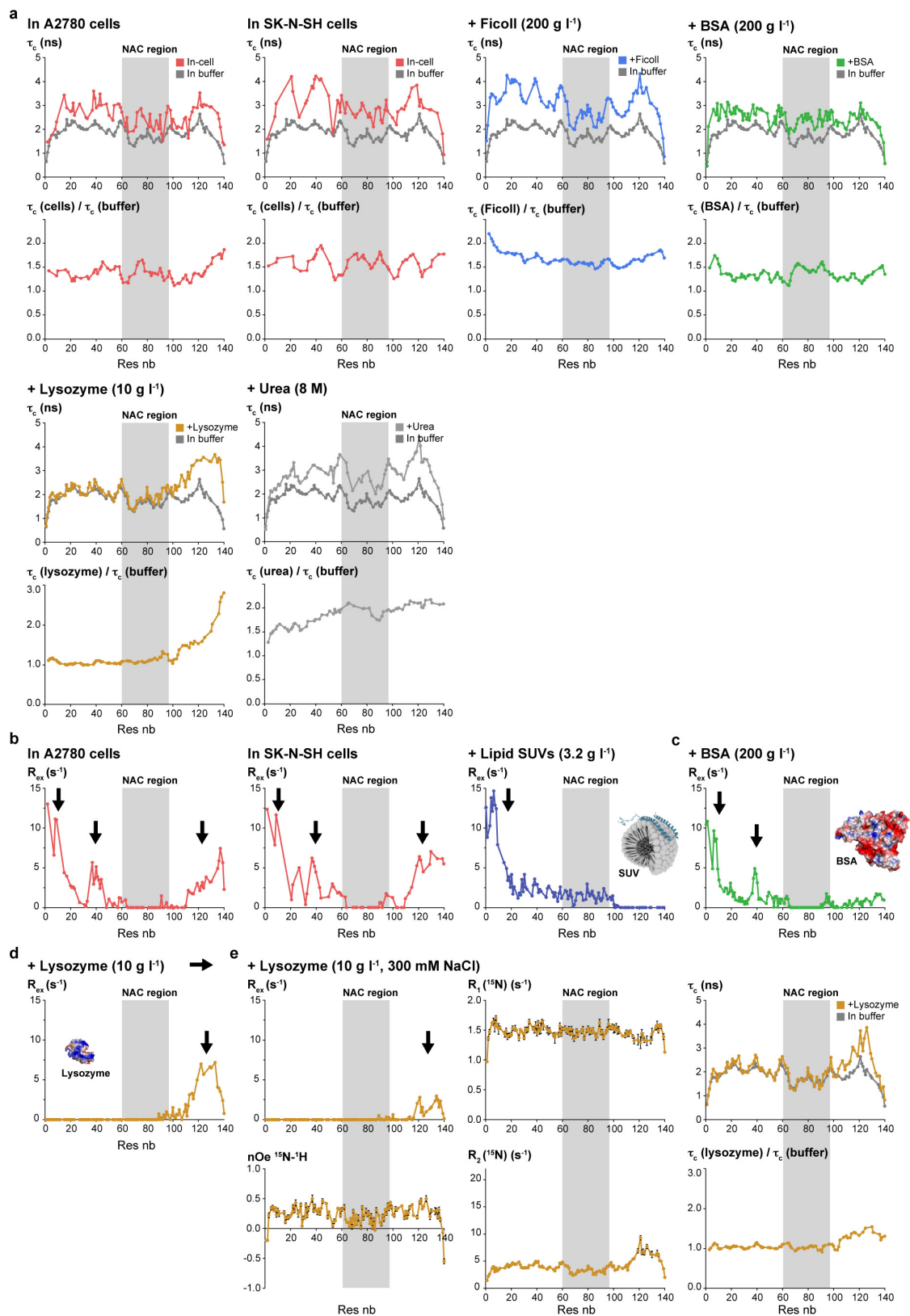
in uniform line broadening of its first  $\sim 100$  residues. **c**, Line-broadening profiles ( $I/I_0$ ) of  $\alpha$ Syn residues in HeLa, RCSN-3 and B65 cells are similar to A2780 and SK-N-SH cells (Fig. 3a), irrespective of different intracellular  $\alpha$ Syn concentrations. For simplification, profiles show values averaged over three consecutively resolved residues. **d**, Representative residue-resolved  $T_1$  relaxation curves acquired on intracellular  $\alpha$ Syn in A2780 and SK-N-SH cells (red), and in buffer (grey). Residues of the NAC are shaded in grey. **e**, Experimental  $T_2$  relaxation curves for residues in different regions of intracellular  $\alpha$ Syn in A2780 (left) and SK-N-SH (right) cells. All in-cell NMR relaxation data were acquired on two independent samples. Measurement points indicate the mean of two independent experiments. Error bars show the range of lowest and highest values.



**Extended Data Figure 5 |  $\alpha$ Syn relaxation rates in A2780 and SK-N-SH cells and in crowded solutions.** Residue-resolved  $^{15}\text{N}$  longitudinal ( $R_1$ ) and transverse ( $R_2$ ) relaxation rates,  $^{15}\text{N}$ - $^1\text{H}$  NOE data, and proton ( $^1\text{H}$ )  $R_2$  values of N-terminally acetylated  $\alpha$ Syn in buffer (grey), A2780 and SK-N-SH cells (red), in the presence of  $200\text{ g l}^{-1}$  Ficoll (blue),  $200\text{ g l}^{-1}$  BSA (green),  $10\text{ g l}^{-1}$  lysozyme (orange), sub-saturating amounts of SUVs

( $3.2\text{ g l}^{-1}$ , purple) and  $8\text{ M}$  urea (grey). All *in vitro* data were acquired on samples at pH 6.4. Relaxation data from measurements at pH 7.4 are available in the respective Source Data. Relaxation measurements were performed on two independent *in vitro* or in-cell NMR samples with data points representing their mean. Error bars show the range of lowest and highest values. Residues of the NAC region are shaded in grey.



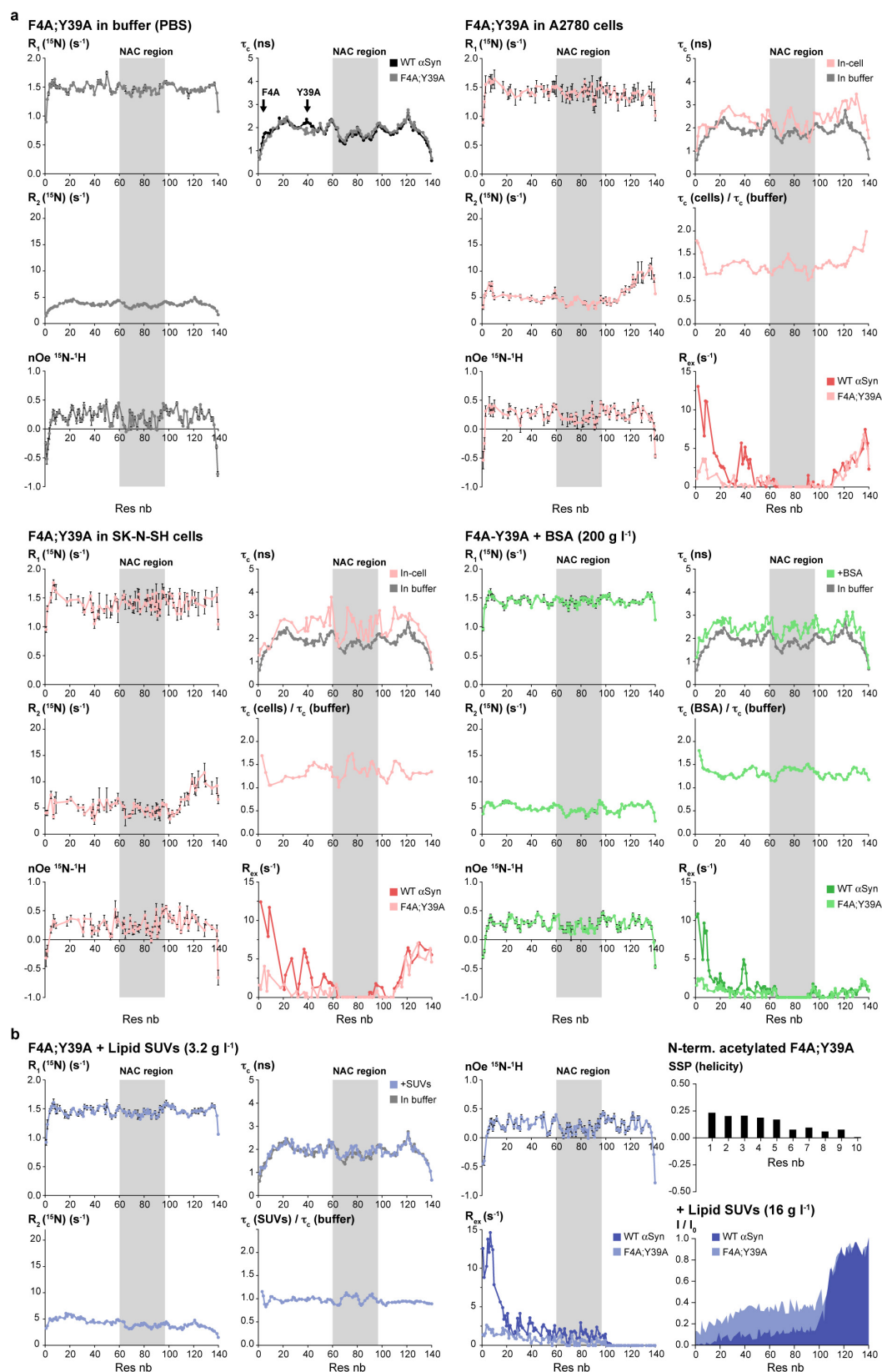


Extended Data Figure 6 | See next page for figure caption.

**Extended Data Figure 6 | Residue-resolved  $\tau_c$  and  $R_{ex}$  profiles of  $\alpha$ Syn in cells and in crowded solutions.** **a**, Rotational correlation times ( $\tau_c$ ) of N-terminally acetylated  $\alpha$ Syn in buffer (grey) and in A2780 and SK-N-SH cells (red). In-cell to in-buffer  $\tau_c$  ratios show average values of three consecutively resolved residues. Overall mobility of intracellular  $\alpha$ Syn is reduced by a factor of  $\sim 1.5$ , which is similar to  $\alpha$ Syn in the presence of 200 g l<sup>-1</sup> glycerol (available in Source Data). Additional mobility profiles and  $\tau_c$  ratios of N-terminally acetylated  $\alpha$ Syn in solutions containing 200 g l<sup>-1</sup> Ficoll (blue), 200 g l<sup>-1</sup> BSA (green), 10 g l<sup>-1</sup> lysozyme (orange) and 8 M urea (light grey) at pH 6.4 (data from measurements at pH 7.4 are available in the respective Source Data). **b**, Calculated  $R_{ex}$  profiles of N-terminally acetylated  $\alpha$ Syn in A2780 and SK-N-SH cells, and in the presence of sub-saturating amounts of SUVs (3.2 g l<sup>-1</sup>, pH 6.4). Regions exhibiting pronounced exchange behaviours are marked with arrows.

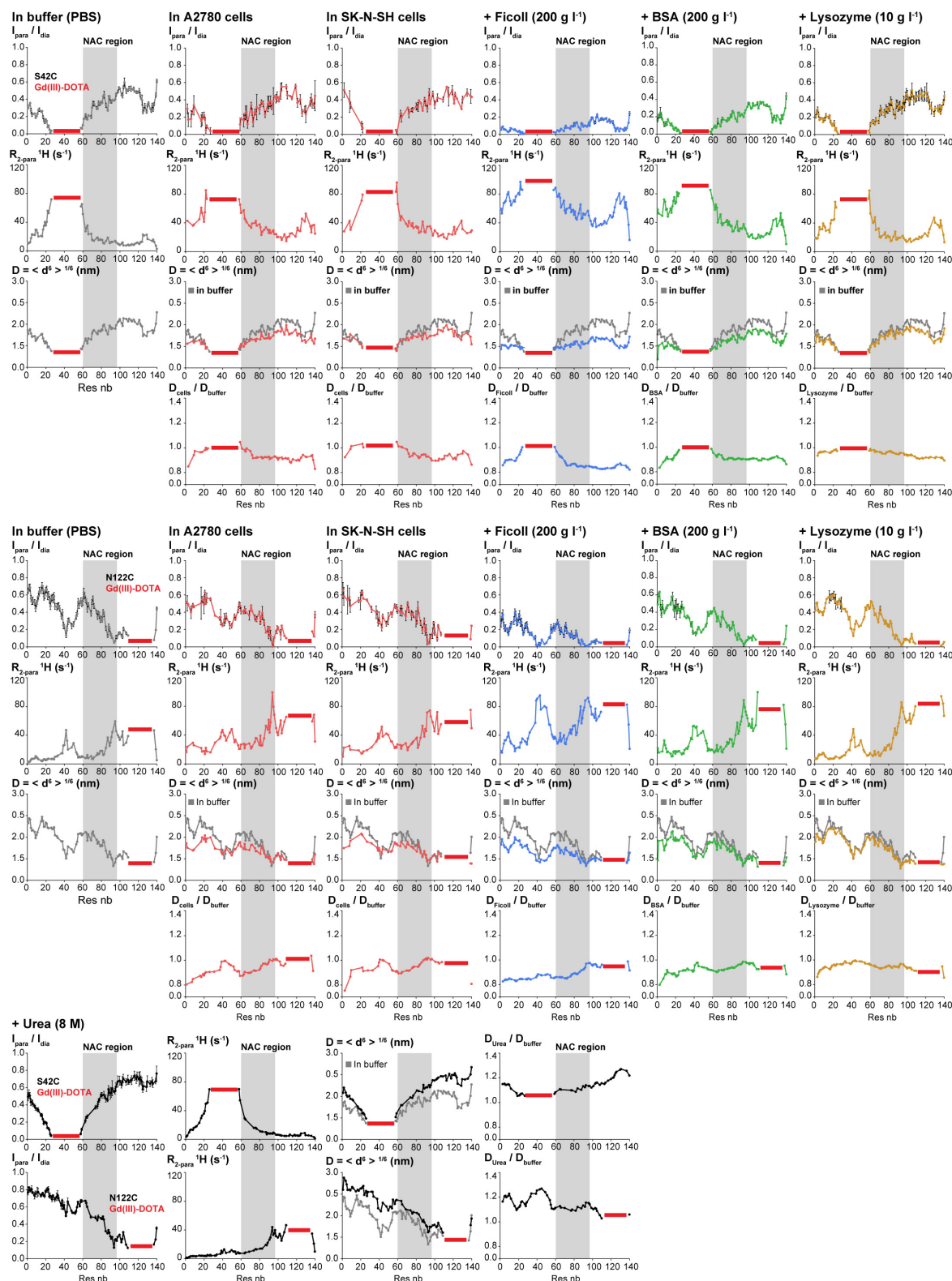
**c, d**,  $R_{ex}$  profiles of N-terminally acetylated  $\alpha$ Syn in the presence of 200 g l<sup>-1</sup> BSA (green) (**c**) and 10 g l<sup>-1</sup> lysozyme (orange) (**d**) at pH 6.4 (data from measurements at pH 7.4 are available in the respective Source Data). Insets depict surface representations of BSA and lysozyme, colour-coded according to their electrostatic surface potentials. Isoelectric points (pI) of BSA and lysozyme are 4.7 and 11.35, respectively. N- and C-terminal line broadening effects observed in cells are recapitulated with SUVs and BSA, and with lysozyme, respectively. **e**, NMR relaxation data,  $\tau_c$  and  $R_{ex}$  profiles of N-terminally acetylated  $\alpha$ Syn in lysozyme-crowded solutions (10 g l<sup>-1</sup>) at 300 mM salt (pH 6.4) reveal diminished exchange contributions for C-terminal  $\alpha$ Syn residues. Data points with error bars represent the mean of two independent experiments. Error bars show the range of lowest and highest values. Residues of the NAC region are shaded in grey.





**Extended Data Figure 7 | Relaxation data of  $\alpha$ Syn(F4A;Y39A) in cells and crowded solutions.** **a**, Residue-resolved relaxation data of N-terminally acetylated  $\alpha$ Syn(F4A;Y39A) in buffer (grey), in A2780 and SK-N-SH cells (light red) and in the presence of 200 g l<sup>-1</sup> of BSA (light green) at pH 6.4 (relaxation data from measurements at pH 7.4 are available in the respective Source Data). **b**, Residue-resolved relaxation data of N-terminally acetylated  $\alpha$ Syn(F4A;Y39A) in the presence of sub-saturating amounts of SUVs (3.2 g l<sup>-1</sup>, pH 6.4).  $R_{ex}$  panels show the comparison between wild-type (WT) and mutant (F4A;Y39A)

$\alpha$ Syn in all cases. All relaxation measurements were performed on two independent samples. Error bars show the range of lowest and highest values. For simplification, profiles of  $\tau_c$  ratios ( $\tau_c/\tau_{c-buffer}$ ) depict values averaged over three consecutively resolved residues. SSP scores indicating residual helicity within the first ten residues of N-terminally acetylated  $\alpha$ Syn(F4A;Y39A) are shown on top of the comparison of signal intensity ratios ( $I/I_0$ ) of wild-type (dark blue) and mutant (light blue) protein in the presence of SUVs.



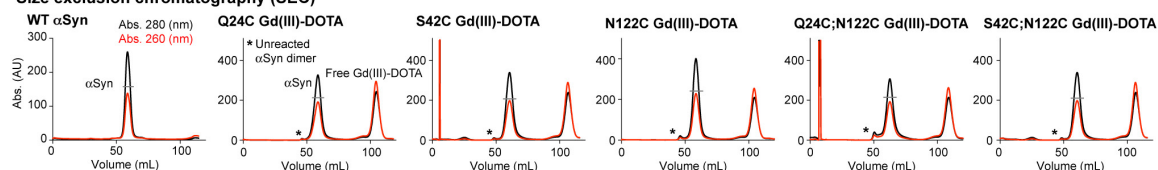
**Extended Data Figure 8 | PRE data of  $\alpha$ Syn in cells and crowded solutions.** Top, PRE profiles of S42C Gd(III)-DOTA tagged N-terminally acetylated  $\alpha$ Syn in buffer, cells and different crowded solutions at pH 6.4 (data obtained at pH 7.4 are available in the respective Source Data). Intensity ratios ( $I_{\text{para}}/I_{\text{dia}}$ ) of  $\alpha$ Syn in mixtures of 70% Gd(III)-DOTA tagged paramagnetic  $\alpha$ Syn and 30% Lu(III)-DOTA tagged diamagnetic  $\alpha$ Syn. 100% diamagnetic Lu(III)-DOTA tagged N-terminally acetylated  $\alpha$ Syn was used to measure reference intensities (that is,  $I_{\text{dia}}$ ). Red bars indicate residues in the vicinity of the Gd(III)-DOTA tag experiencing proximal PRE effects and line broadening beyond the detection limit. Measurement points with error bars indicate the mean of two independent experiments.

Error bars show the range of lowest and highest values. Paramagnetic  $R_2$  ( $^1\text{H}$ ) rates and calculated PRE distance averages ( $\langle d^6 \rangle^{1/6}$ ) in comparison to values measured in buffer (grey) are indicated. PRE-derived distances were obtained assuming that every Gd(III)- $^1\text{H}$  vector fluctuation rate scales linearly with that of the backbone N-H vector and the Gd(III)-DOTA complex. Distance ratios ( $D_{\text{cell/crowding agent}}/D_{\text{buffer}}$ ) indicate levels of  $\alpha$ Syn compaction and represent average ratios over three consecutively resolved residues. Middle, same PRE experiments with N122C Gd(III)-DOTA tagged, N-terminally acetylated  $\alpha$ Syn. Bottom, PRE experiments with S42C and N122C Gd(III)-DOTA tagged, N-terminally acetylated  $\alpha$ Syn in the presence of 8 M urea.

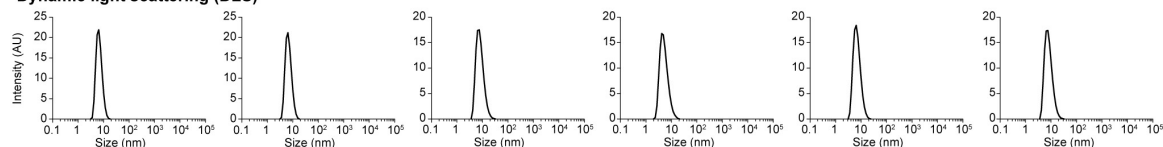


a

## Size exclusion chromatography (SEC)

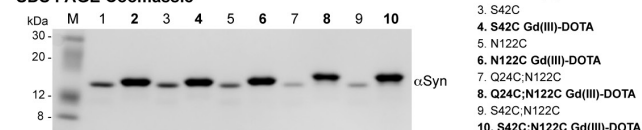


## Dynamic light scattering (DLS)

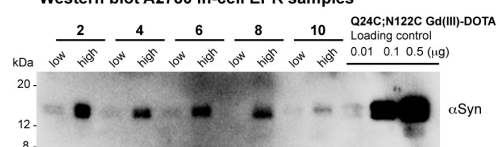


b

## SDS PAGE Coomassie

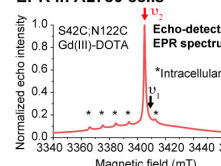


## Western blot A2780 in-cell EPR samples

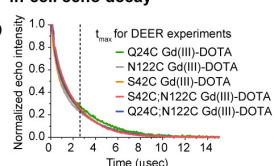


c

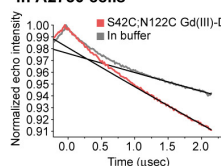
## EPR in A2780 cells



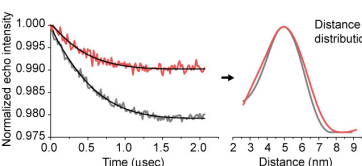
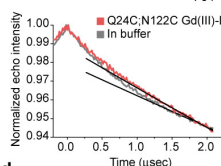
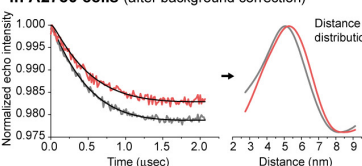
## In-cell echo decay



## In A2780 cells

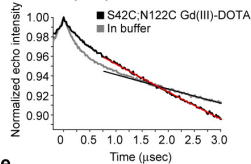


## In A2780 cells (after background correction)



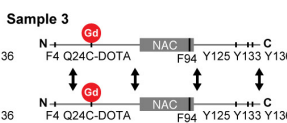
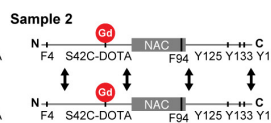
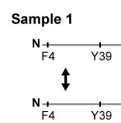
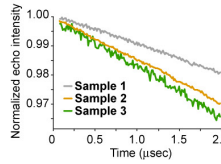
d

## + Urea (8 M)



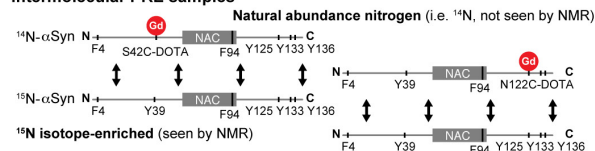
e

## In A2780 cells (intermolecular DEER)

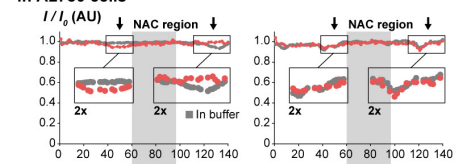
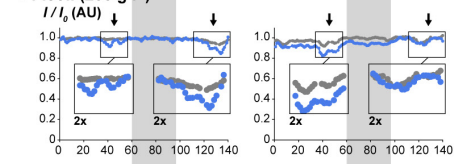
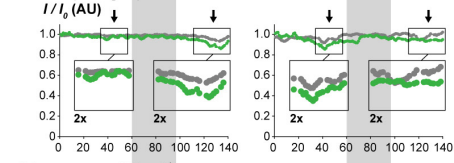
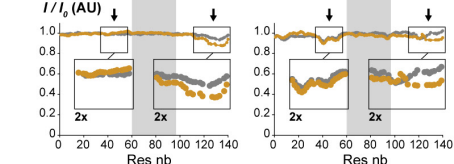


f

## Intermolecular PRE samples



## In A2780 cells

+ Ficoll (200 g l<sup>-1</sup>)+ BSA (200 g l<sup>-1</sup>)+ Lysozyme (10 g l<sup>-1</sup>)

Extended Data Figure 9 | See next page for figure caption.

**Extended Data Figure 9 | Stable Gd(III)-DOTA tagged  $\alpha$ Syn for in-cell PRE and EPR-DEER experiments.** **a**, Size-exclusion chromatography (SEC) profiles of single- (that is, Q24C, S42C and N122C) and double- (that is, Q24C;N122C and S42C;N122C) Gd(III)/Lu(III)-DOTA-tagged  $\alpha$ Syn (top) and corresponding dynamic light scattering (DLS) of monomeric protein fractions (bottom). **b**, SDS-PAGE separation of Gd(III)-DOTA  $\alpha$ Syn reveals characteristic molecular mass differences of single- and double-tagged samples (left). Western blot of exemplary A2780 in-cell EPR samples ( $20 \times 10^6$  cells) electroporated with 125  $\mu$ M (low) or 250  $\mu$ M (high) Gd(III)-DOTA  $\alpha$ Syn (right). For gel source data, see Supplementary Fig. 1. **c**, Echo-detected in-cell EPR spectrum of S42C-N122C Gd(III)-DOTA  $\alpha$ Syn in A2780 cells (left). EPR signals of intracellular Mn(II) are indicated with asterisks. Red and black arrows denote microwave pump ( $\nu_2$ ) and observe ( $\nu_1$ ) frequencies, respectively. In-cell echo decays of the single- and double- Gd(III)-DOTA tagged  $\alpha$ Syn samples are shown on the right. DEER traces of S42C;N122C and Q24C;N122C Gd(III)-DOTA  $\alpha$ Syn in A2780 cells (red), and in buffer (grey) after background correction. Calculated distance distributions are shown on the right. **d**, DEER traces of S42C;N122C Gd(III)-DOTA  $\alpha$ Syn in the presence of 8 M urea (black) and in buffer (grey). Data were recorded on a different protein batch sample to that shown in **c**.

**e**, Intermolecular DEER measurements of three A2780 in-cell EPR samples carrying single Gd(III)-DOTA  $\alpha$ Syn (that is, N122C, S42C and Q24C) at intracellular concentrations of  $40 \pm 10 \mu$ M reveal absence of detectable spin-spin coupling, as evident from straight decays of the corresponding DEER traces. Cartoon illustrations on the right depict possible  $\alpha$ Syn- $\alpha$ Syn contacts with orientations chosen arbitrarily. **f**, Intermolecular PRE profiles of mixtures (50:50) of  $^{15}$ N isotope-enriched  $\alpha$ Syn and of Gd(III)-DOTA (S42C or N122C)  $\alpha$ Syn containing natural abundance nitrogen in A2780 cells (red, intracellular  $\alpha$ Syn concentrations  $\sim 100 \mu$ M), and in buffer (grey). Cartoon illustrations depict possible  $\alpha$ Syn- $\alpha$ Syn contacts with orientations chosen arbitrarily. Intermolecular PRE effects are manifested by selective signal attenuations ( $I/I_0$ ) of the isotope-enriched, NMR-visible portion of the  $\alpha$ Syn mixture ( $I$ ) over reference samples containing only  $^{15}$ N isotope-enriched  $\alpha$ Syn ( $I_0$ ). For simplification,  $I/I_0$  profiles show values averaged over three consecutively resolved residues. Experiments with reconstituted mixtures of  $\alpha$ Syn (100  $\mu$ M) in the presence of Ficoll (blue, 200 g l $^{-1}$ ), BSA (green, 200 g l $^{-1}$ ) and lysozyme (orange, 10 g l $^{-1}$ ) are shown below. Insets depict magnified regions of preferred  $\alpha$ Syn- $\alpha$ Syn contacts. Intermolecular PRE effects are largest for solutions containing Ficoll and BSA, whereas they are not detected in A2780 cells. All  $\alpha$ Syn constructs are N-terminally acetylated.

# Persistent HIV-1 replication maintains the tissue reservoir during therapy

Ramon Lorenzo-Redondo<sup>1\*</sup>, Helen R. Fryer<sup>2\*</sup>, Trevor Bedford<sup>3</sup>, Eun-Young Kim<sup>1</sup>, John Archer<sup>4</sup>, Sergei L. Kosakovsky Pond<sup>5†</sup>, Yoon-Seok Chung<sup>6</sup>, Sudhir Penugonda<sup>1</sup>, Jeffrey G. Chipman<sup>7</sup>, Courtney V. Fletcher<sup>8</sup>, Timothy W. Schacker<sup>9</sup>, Michael H. Malim<sup>10</sup>, Andrew Rambaut<sup>11</sup>, Ashley T. Haase<sup>12</sup>, Angela R. McLean<sup>2</sup> & Steven M. Wolinsky<sup>1</sup>

**Lymphoid tissue is a key reservoir established by HIV-1 during acute infection. It is a site associated with viral production, storage of viral particles in immune complexes, and viral persistence. Although combinations of antiretroviral drugs usually suppress viral replication and reduce viral RNA to undetectable levels in blood, it is unclear whether treatment fully suppresses viral replication in lymphoid tissue reservoirs. Here we show that virus evolution and trafficking between tissue compartments continues in patients with undetectable levels of virus in their bloodstream. We present a spatial and dynamic model of persistent viral replication and spread that indicates why the development of drug resistance is not a foregone conclusion under conditions in which drug concentrations are insufficient to completely block virus replication. These data provide new insights into the evolutionary and infection dynamics of the virus population within the host, revealing that HIV-1 can continue to replicate and replenish the viral reservoir despite potent antiretroviral therapy.**

Combinations of antiretroviral drugs routinely impair HIV-1 production and replication to levels that are undetectable in the blood within weeks of starting treatment<sup>1</sup>. None of the current treatments, however, is capable of eradicating the virus from a long-lived reservoir in resting memory CD4<sup>+</sup> T cells and other cell types that potentially protect the virus from antiretroviral drugs or immune surveillance<sup>2–5</sup>. Intermittent virus production from reactivation of a small proportion of latently infected CD4<sup>+</sup> T cells (rather than low levels of ongoing replication) is thought to drive viral rebound detected in blood of treated patients with well-suppressed infection<sup>6–8</sup>. Ongoing replication is considered unlikely because neither viral genetic divergence over time, nor the emergence of drug resistance mutations have been convincingly documented<sup>9,10</sup>. As earlier studies only examined viral sequences derived from the blood of patients who continued to suppress viral replication in that anatomic compartment<sup>11</sup>, the conclusions are not necessarily generalizable to other compartments in the body, particularly to lymphoid tissue where the frequency of infection per cell is mostly higher<sup>12</sup> and the intracellular drug concentrations are much lower than in blood<sup>13</sup>. Under low drug concentrations, the virus may continue to replicate and evolve in ‘sanctuary sites’ within the reservoir of cells in lymphoid tissue, and remain undetectable in the bloodstream for a time depending on viral population migration dynamics between the two compartments. Here we use a multi-pronged strategy of deep-sequencing, time-calibrated phylogenetic analysis, and mathematical modelling to characterize the distinct temporal structure and divergence of compartmentally sampled viral sequences. We discover ongoing replication in lymphoid tissue sanctuary sites of patients despite undetectable blood levels of virus. Our sampling approach differs fundamentally from those of previous studies<sup>14–16</sup>, which do not address evolutionary dynamics within lymphoid tissue, and better suits investigation of the

dynamic nature of the viral reservoir during treatment with potent antiretroviral drugs.

## HIV-1 sequence determination

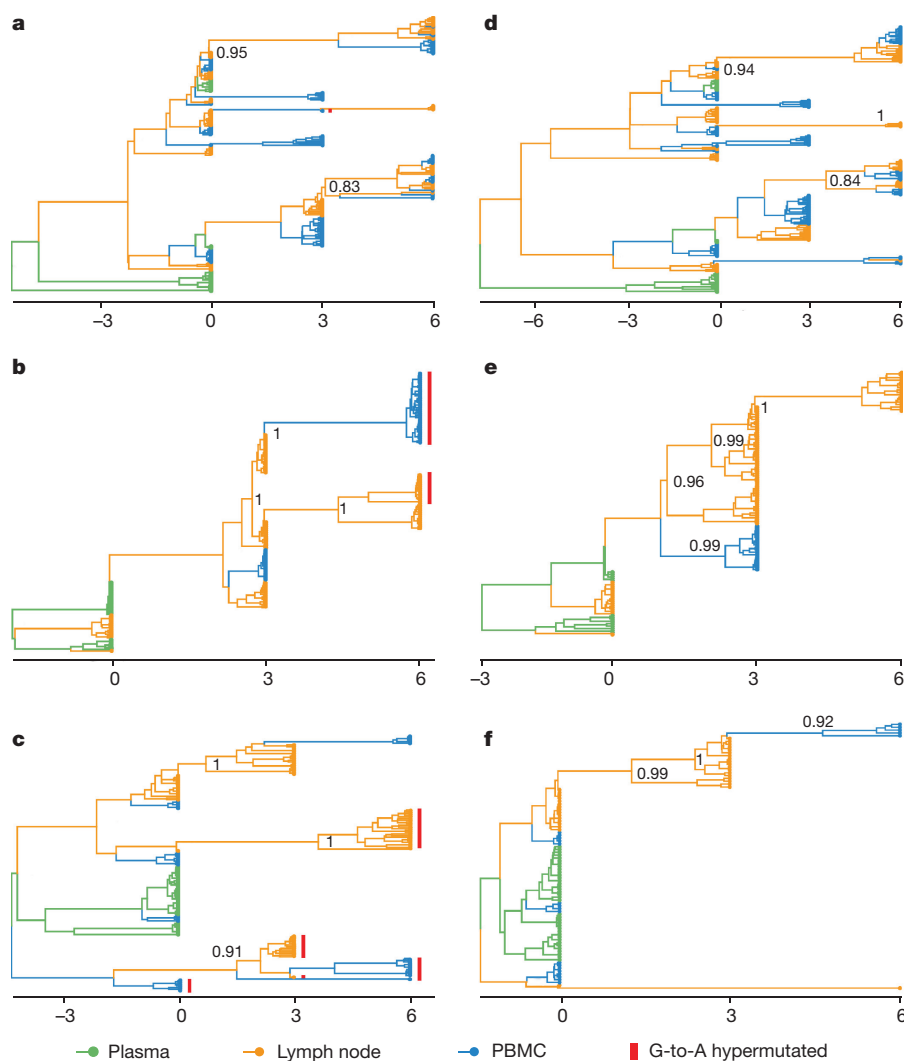
To investigate the evolution and spatial dispersion of virus with high accuracy, we deep sequenced (using the GS FLX sequencing platform (Roche)) HIV-1 DNA in cells from blood and inguinal lymph nodes collected from three subjects at three separate times (at day 0, and after 3 and 6 months of treatment), described elsewhere<sup>13</sup>. Previous work established that viral sequences contemporaneously sampled from lymphoid tissue in different locations are genetically homogeneous<sup>17</sup>, consistent with CD4<sup>+</sup> T cell homing and trafficking<sup>18</sup>. Consequently, detailed assessment of a portion of a solitary lymph node is no more susceptible to bias than wider anatomical sampling. We also sequenced viral RNA in the plasma (day 0) from these three study subjects. Two subjects (1727 and 1679) had well-suppressed infections (<48 viral genome copies per ml of plasma); and the third subject (1774) continued to have measureable amounts of viral RNA in plasma after 3, but not 6, months of treatment (Extended Data Fig. 1). Subjects 1727 and 1679 were each infected with HIV-1 for approximately 3 to 4 months and were antiretroviral-drug-naïve before the study. Subject 1774 was infected with HIV-1 for approximately 17 years and was antiretroviral-therapy-experienced, but had not received any treatment for at least 1 year before the study.

We aligned individual reads with an average length of 548 basepairs (bp) to a consensus viral sequence for each patient using reference-guided assembly, and corrected sequencing errors for potentially inflated estimates of genetic diversity<sup>19</sup>. We then used a previously described approach<sup>20</sup> to reconstruct the minimum number of viral haplotypes needed to adequately explain the observed reads. We

<sup>1</sup>Division of Infectious Diseases, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60011, USA. <sup>2</sup>Institute for Emerging Infections, Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK. <sup>3</sup>Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. <sup>4</sup>Centro de Investigação em Biodiversidade e Recursos Genéticos Universidade do Porto, 4485-661 Vairão, Portugal. <sup>5</sup>Department of Medicine, University of California, San Diego, California 92093, USA. <sup>6</sup>Division of AIDS, Center for Immunology and Pathology, Korea National Institutes of Health, Chungju-si, Chungcheongbuk-do, 28159, South Korea. <sup>7</sup>Department of Surgery, University of Minnesota, Minneapolis, Minnesota 55455, USA. <sup>8</sup>Antiviral Pharmacology Laboratory, University of Nebraska Medical Center, College of Pharmacy, Omaha, Nebraska 68198, USA. <sup>9</sup>Division of Infectious Diseases, University of Minnesota, Minneapolis, Minnesota 55455, USA. <sup>10</sup>Department of Infectious Diseases, King's College London, Guy's Hospital, London SE21 7DN, UK. <sup>11</sup>Centre for Immunology, Infection and Evolution, University of Edinburgh, Edinburgh EH9 3FL, UK. <sup>12</sup>Department of Microbiology, University of Minnesota, Minneapolis, Minnesota 55455, USA. <sup>†</sup>Present address: Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, Pennsylvania 19122-1801, USA.

\*These authors contributed equally to this work.





**Figure 1 | Time-structured phylogenetic history of haplotypes in lymph nodes and blood.** a–f, MCC phylogenetic trees constructed from the complete alignments of the haplotypes from the Gag region of HIV-1 for subjects 1774, 1727, and 1679 with all haplotypes (a, b, and c, respectively) and with the haplotypes containing G-to-A hypermutations removed (d, e, and f, respectively). Branch colours represent the most

probable (modal) anatomic location of their descendent node inferred through Bayesian reconstruction of the ancestral state, along with the posterior probabilities of the most probable location of the major nodes. The abscissa on the plots represents the time in months from the initiation of therapy. PBMC, peripheral blood mononuclear cells.

calculated the sequencing error rate and set the cut-off for the subsequent analyses using a known internal control sequence. We found no significant evidence for recombinant sequences that could bias the analysis. High coverage enabled us to correct PCR and sequencing errors (leaving an average 25,000 final long-reads per sample) to detect variants present in at least 0.04% of the virus population (see Methods). The sequences from the Pol region of HIV-1 that spanned the genomic region encoding the viral enzymes protease or reverse transcriptase showed no evidence of new mutations that confer resistance to the particular antiretroviral drug used (data not shown).

To avoid uncertainties in the haplotype detection due to sparse sampling and prevent systematically biased evolutionary analyses, we established a higher number of template molecules than the depth of the sequence data by a median 9.6-fold (range 3.2- to 362-fold). The high coverage of ultra-deep sequencing ensured reliable detection of low-frequency viral variants (see Methods). In support of this conclusion, we found limited variability across inferred haplotypes in two completely independent technical replicates made from the same RNA or DNA sample at each time point from each subject using the same procedures (average Spearman rank correlation coefficient between single nucleotide polymorphism frequencies across replicates of  $\rho = 0.832$ , interquartile range across samples,  $\rho = 0.820$ – $0.851$ ; 93.7% of haplotypes above 1%

frequency appeared above 1% frequency in the replicate; 97.7% of haplotypes below 1% frequency appeared below 1% frequency in the replicate), indicating that the haplotype representation is not notably biased from random amplification of some sequences and not others. The high degree of concordance between technical replicates validated our approach for the computational characterization of the viral populations, which is robust with respect to experimental error and stochastic effect<sup>20</sup>.

### Phylogenies show temporal structure

The inferred haplotypes corresponding to the Gag or Pol regions of HIV-1 were subjected to maximum-likelihood methods of phylogeny estimation (Extended Data Figs 2 and 3). We masked out the guanine editing sites that fall within trinucleotide contexts targeted by cellular APOBEC3 proteins APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H from the alignments to avoid distortion and retain the phylogenetic information<sup>20</sup> (Extended Data Figs 2 and 3). Phylogenetic relationships between the distinct haplotypes showed a temporal structure consistent with the molecular clock (continuing nucleotide substitutions occurring at a constant rate), as evidenced by strong correlation between root-to-tip distance and sampling date in the regression analyses (despite the short branches). Branch support computed using an approximate likelihood ratio test and the proportion of sites that

are different ( $p$ -distance) verify the divergence of haplotypes between day 0 and after 6 months of treatment in most of the Gag and Pol regions of HIV-1 analysed (Extended Data Table 1). Consistent with ongoing replication rather than sampling of different virus populations in lymph nodes, viral sequences contemporaneously sampled from lymphoid tissue and blood showed a similar degree of divergence. With removal of the haplotypes that are found to harbour repetitive inactivating base substitutions of guanosine-to-adenosine (G-to-A), the evolutionary lineage emerged that lead up to the APOBEC3-mediated hypermutation event, and the evolutionary rate estimates (range,  $6.24 \times 10^{-4}$  to  $1.02 \times 10^{-3}$  substitutions per site per month; Extended Data Table 2) are consistent with those of intra-host virus estimations (range,  $5.22 \times 10^{-4}$  to  $8.42 \times 10^{-4}$  substitutions per site per month)<sup>21</sup>. We therefore conclude that continued virus replication contributes to the viral reservoir.

### Viral populations show compartmentalized structure

The pairwise fixation index ( $F_{ST}$ ), a standard measure of genetic differentiation between populations, and the nearest-neighbour statistic ( $S_{nn}$ ), a measure of spatial relationships in populations, confirmed substantial genetic variation between lymph node and blood at each time point (Extended Data Table 3)<sup>22,23</sup>. As sequence evolution in spatially distinct compartments may be affected by selection differentially, we used an unrestricted branch-site random effects model to test whether the proportion of sites along a subset of tree branches subject to episodic diversifying selection is significantly different from zero<sup>24</sup> (see Methods and Extended Data Table 4). We restricted the test to internal branches, which capture at least one, but probably multiple, rounds of virus replication to reduce the biasing effects of neutral or deleterious mutations on the ratio of nonsynonymous-to-synonymous substitution rate ( $\omega$ ) estimates where selection has not yet fully filtered such population-level variation<sup>25,26</sup>. Except for one study subject (1679), where a small proportion of sites (0.3%) were under strong diversifying positive selection along internal tree branches (likelihood ratio test,  $P \approx 10^{-6}$ ), we found little evidence to suggest that the virus is evolving in response to strong selective forces. We concluded that there is only restricted migration of haplotypes from one compartment to another.

### The phyloanatomic history of HIV-1

To infer evolutionary patterns and population-dynamic processes from the time-structured sequence data, we used a Bayesian statistical framework that estimates the substitution rate, divergence time, and demographic history of the sampled viral lineages to structure-rooted time-resolved phylogenies<sup>27,28</sup> (see Methods). This approach resolves evolutionary patterns to infer the timing and direction of the key migrations of the virus within hosts. Figure 1 shows the phyloanatomic history of HIV-1 within the study subjects inferred from the ancestral and descendent haplotypes on a Bayesian maximum clade credibility (MCC) phylogenetic tree. The branching patterns in the trees, which reconstruct the origins and trace the flow of HIV-1 within hosts, and the tissue of origin of the internal nodes show strong statistical support (highlighted by their posterior probabilities). Branch lengths in these time-structured trees (left panels with all haplotypes; right panels with the putative G-to-A hypermutated sequences removed to avoid distortion) represent posterior median estimates of calendar time. The temporal structure in the trees shows a strong clock-like signal and rates consistent with HIV-1 within-host evolution. The best-fit model included a strict molecular clock and assumed a constant population size, though a model with a relaxed molecular clock gave qualitatively similar results (data not shown).

Based on inferred MCC tree topologies and the compartment assignments to unobserved internal nodes, an underlying conformity and strong correlation exist between the genetic and anatomic locations and the spread direction of the virus in the body. A particular pattern recurs: the haplotypes in lymph nodes are the source of viral lineages that migrate from lymph node to blood (Fig. 1). We deduce that viral lineages in blood are derived from replicating virus in lymph

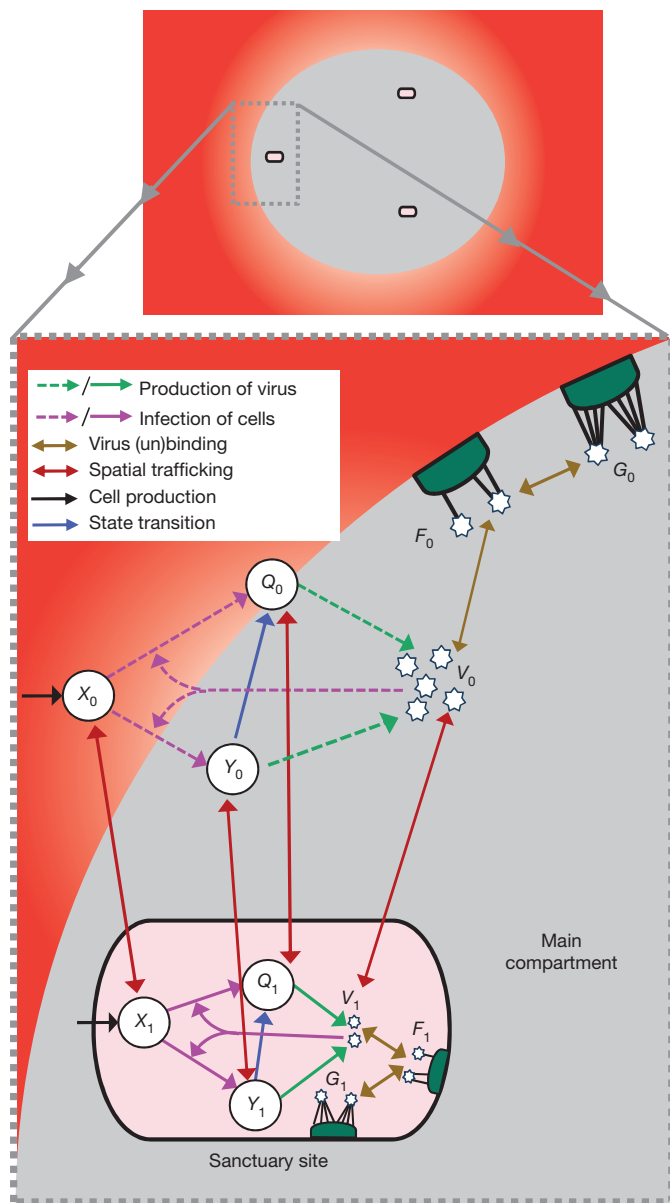
nodes with little or no evidence of an additional source in blood. The time-scaled trees show a strong and significant result and are robust to different substitution models (see Methods). These data, only revealed through temporarily and spatially resolved sampling, further support the conclusion that the pattern does not result from distinct populations of haplotypes being sampled from different compartments, but rather migration and colonization of haplotypes between lymphoid tissue and blood. A structured coalescent model<sup>29</sup>, less prone to potential bias in spatial inference estimates, shows higher migration rates from lymph node to blood (Extended Data Table 5), confirming that the direction of flow is not due to oversampling of a particular anatomic location that would have increased estimates of traffic into that location<sup>30</sup>.

Our results, which reconstruct the dynamics of HIV-1 spread within the body, imply that in patients with no detectable viral RNA in plasma, the virus reservoir is constantly replenished by low-level virus replication in lymphoid tissue. Distinguishing between low amounts of viral replication and pools of latently infected cells that may persist and reactivate HIV-1 infection is methodologically difficult. A small number of HIV-1 sequences isolated at consecutive time points that persisted without evidence of genetic change might result from long-lived central memory cells, a fraction of which may have reverted to a resting state, or latently infected transitional memory cells that persist by clonal expansion (driven by homeostatic proliferation) or survival of long-lived infected CD4<sup>+</sup> T cells that contain replication-competent virus<sup>8,15,22,31–34</sup>. Two of the subjects (1774 and 1679) showed that some haplotypes persist as a single tree branch through time (Fig. 1d and f), consistent with proliferation of HIV-1-infected cells or long-term cell survival<sup>33,34</sup>. Regardless of the different mechanisms for self-renewal/persistence by which some of these quite similar latent or defective viral lineages may have persisted, these quiescent viruses differ from others that have evolved and trafficked between compartments. The temporally and compartmentally sampled data show that viral lineages continue to diverge in well-suppressed patients and help to explain the persistence of infectious viral reservoirs with little reduction of the virus pool<sup>35</sup>. The dynamic nature of the viral population in lymphoid tissue sanctuaries—where infected cells can still produce new viruses, infect new target cells, and replenish the pool—undermines previous estimates of the time necessary to purge the reservoir of latently infected cells and achieve virus eradication<sup>3</sup>.

### A spatial and dynamic model

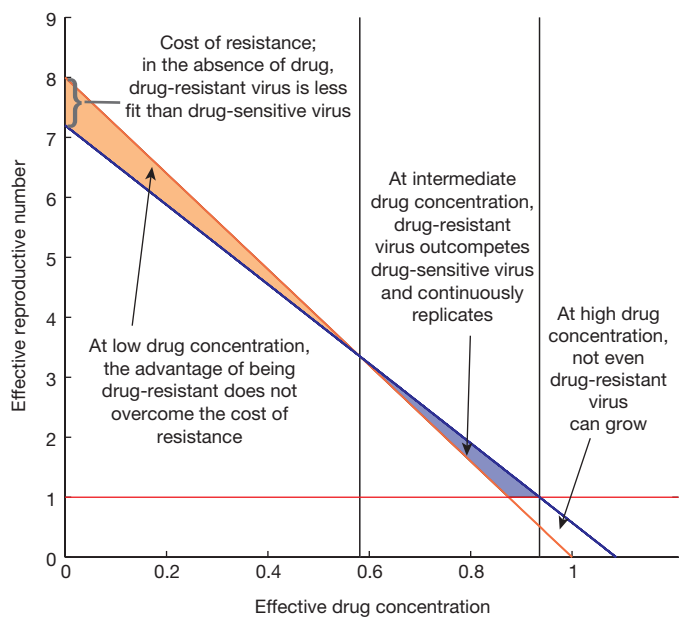
Although viral genetic diversity accumulated over time, the infected cells did not produce new virus with drug-resistance mutations, conferring a putative fitness advantage that might have led to a systemic viral rebound. To provide a mechanistic explanation of this scenario, we developed a drug-concentration-dependent mathematical model of virus replication and spread between spatially distinct compartments to explore the deterministic components of viral dynamics<sup>36</sup>. In this model (see Fig. 2 and Supplementary Information), the virus can occupy two spatially distinct compartments that have limited traffic of virus particles or cells. The main compartment has a larger volume and high effective drug concentration. The other, smaller volume compartment (<0.01% of the size of the main compartment) has a lower drug concentration and represents a sanctuary site within the lymphoid tissue reservoir. The model includes competition between two viral strains<sup>37</sup>, one that is sensitive to a potent multidrug regimen and one that is partially resistant to the full complement of antiretroviral drugs, but less fit than the drug-sensitive strain in the absence of treatment. Within each compartment the balance between the strains is determined by the difference in their replicative fitness without treatment, the effectiveness of drug therapy to curb new rounds of infection in that region (determined by drug concentration), and the susceptibility of each strain to antiretroviral therapy<sup>38</sup>.

Figure 3 illustrates a hypothetical fitness landscape, which portrays the relationship between the fitness of each strain and the evolutionary adaptation across a range of drug concentrations. The fitness landscape



**Figure 2 | Cartoon illustration of the drug concentration-dependent spatial model.** In the main compartment ( $i = 0$ ; the majority of lymphoid tissue and the blood), drug concentration is high (grey and red). In the sanctuary site ( $i = 1$ ; a small fraction of the lymphoid tissue and localized extracellular fluid), drug concentration is low (pink). There are uninfected cells ( $X_i$ ), long-lived infected cells ( $Y_i$ ), and short-lived infected cells ( $Q_i$ ), as well as virus particles ( $V_i$ ) that can be bound by few ( $F_i$ ) or many ( $G_i$ ) receptors on the follicular dendritic cell network. The dashed lines represent the effect of treatment in blocking infection and production of infectious virus particles. For graphical simplicity, we do not show the emergence of drug resistance, the production of non-infectious virus particles, virus clearance, nor cell death.

explains the persistence of the pool of virus that is unaffected by antiretroviral drugs. The effective reproductive number  $R$  (the average number of secondary *de novo* infections of cells produced by one infected cell) is a function of the basic reproductive number  $R_0$  and the effectiveness of treatment ( $R = R_0$  in the absence of treatment). The fitness cost of drug resistance determines the difference between the effective reproductive numbers for drug-sensitive and drug-resistant strains ( $R_S$  and  $R_R$ , respectively) at zero effective drug concentration. In a competitive system where both strains are present, the maximum of the two effective reproductive numbers  $R_S$  and  $R_R$  equals the effective reproductive number  $R$  for the system as a whole at that effective drug concentration.

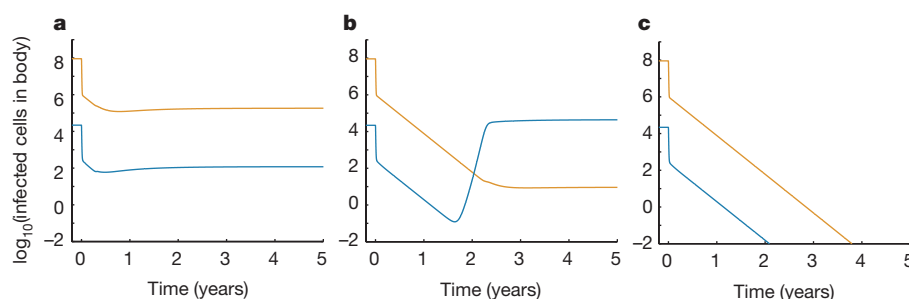


**Figure 3 | Drug-dependent fitness landscape.** Effective reproductive numbers for drug-sensitive ( $R_S$ , orange line) and partially drug-resistant ( $R_R$ , blue line) strains are driven by the effective drug concentration in the relevant compartment. Black vertical lines mark thresholds separating three possible outcomes. When effective drug concentrations are low, the benefit of drug resistance does not overcome the fitness cost of mutations and drug-sensitive strains dominate. Virus replication is sufficient to support generation of partially drug-resistant strains, but these strains will be outcompeted to low numbers and the waiting time to generate a highly drug-resistant strain will be very long (Supplementary Information). At intermediate effective drug concentrations, however, partially drug-resistant strains will dominate. At high concentrations, both  $R_S$  and  $R_R$  fall below one (red line), neither strain can grow and virus replication is halted.

With two spatial compartments, heterogeneity in the distribution of the drug can lead to heterogeneity in  $R$ . In the sanctuary site, where drug penetration is low, the drug-selective pressure on the replicating virus population is too low to compensate for the fitness costs associated with resistance; thus, the effective reproductive number for a drug-sensitive strain is greater than that of the partially drug-resistant strain ( $R_S > R_R$ ), enabling *de novo* infection dominated by the drug-sensitive strain. As drug effectiveness increases, the partially drug-resistant strain gradually becomes more fit relative to the drug-sensitive strain such that at some intermediate drug concentrations, characterized by threshold levels, the effective reproductive number for the resistant strain can be greater than that for the drug-sensitive strain ( $R_R > R_S$ ). This would allow for *de novo* infection dominated by the partially drug-resistant strain. In our model, however, we assume that drug concentration in the main compartment exceeds this threshold. At this high drug concentration, infection is no longer sustainable because the effective reproductive numbers ( $R_R$  and  $R_S$ ) are less than one for both strains.

By assuming a relatively simple two-compartment model with drug concentration differences between them the model predicts that virus, dominated by the drug-sensitive strain, can proliferate in a small sanctuary site where the drug concentration is low (Fig. 4a). Although all single point mutations will be generated sufficiently often to prompt partially drug-resistant strains within individuals, their numbers will remain low through competition. The likelihood is exceedingly low (see Supplementary Information and Supplementary Table 2) of step-wise accumulation of mutations from a partially drug-resistant strain to one that confers resistance to all drugs, each of which would have to come to fixation in the absence of drug selection (Extended Data Fig. 4), or the presence or absence of recombination. The model calculations show that increasing drug effectiveness or penetration across spatial regions can affect evolutionary dynamics, and lead either to





**Figure 4 | Modelling replication dynamics at different effective drug concentrations.** Changes in the total number of cells infected with drug-sensitive (orange line) or partially drug-resistant (blue line) strains in the body during therapy are shown. In the main compartment, combinations of drugs that inhibit viral reverse transcriptase and protease control infection with drug-sensitive virus. The effective drug concentrations of the reverse transcriptase inhibitor ( $z_0$ ) and the protease inhibitor ( $\tilde{z}_0$ ) in this compartment are therefore equal to 1 ( $z_0 = \tilde{z}_0 = 1$ ). In the sanctuary site, we model three different scenarios. **a**, The effective drug

concentrations of the two drug classes in the sanctuary ( $z_1$  and  $\tilde{z}_1$ ) are low ( $z_1 = \tilde{z}_1 = 0.3$ ) favouring drug-sensitive strains. Here, partially drug-resistant strains exist, but at levels below those favouring stepwise evolution towards a fully drug-resistant strain. **b**, Intermediate effective drug concentrations ( $z_1 = \tilde{z}_1 = 0.6$ ) favour partially drug-resistant virus at levels that may be sufficient for evolution towards a fully drug-resistant strain. **c**, High effective drug concentrations ( $z_1 = \tilde{z}_1 = 1$ ) favour the decline of all strains with the cessation of virus replication.

the emergence of drug-resistant strains (Fig. 4b) or the elimination of ongoing replication (Fig. 4c). Our model predictions fit the data well (Extended Data Fig. 5) and confirm that both competition between strains and regional spatial heterogeneity in antiretroviral concentrations help capture the observed dynamics of the viral reservoir in these well-suppressed patients.

Although probabilistic models suggest (and this model allows) production of partially drug-resistant strains, any which arise cannot become dominant in the sanctuary site because of low drug penetration and competition from drug-sensitive strains. Equally, they cannot repopulate the larger, main compartment where high drug concentrations preclude any ongoing replication. In agreement with our phylogenetic inferences, these results suggest that the low-level viral replication in lymphoid tissues where the concentration of antiretroviral drugs is low could allow drug-sensitive strains to grow and spillover to reach the blood<sup>39</sup>.

## Conclusion

Our results reveal how dynamic and spatial processes act together to permit HIV-1 to persist within the infected host and avoid development of resistance despite antiretroviral therapy. From these temporally and compartmentally structured sequence data, we conclude that continued virus production from infected cells in lymphoid tissue sanctuary sites, where drug concentrations are not fully suppressive, can continue to replenish the viral reservoir and traffic to blood or lymphoid tissue<sup>18</sup>. We further show that the virus does not inevitably develop resistance to antiretroviral drugs because the lower concentration of drugs in the sanctuary sites is not sufficient to confer a competitive advantage upon drug-resistant strains. Our findings explain the failure of treatment intensification to fully suppress *de novo* infection and highlight issues surrounding the barriers to delivering antiretroviral drugs at clinically effective concentrations in the infectious viral reservoir. The state-of-the-art sequencing approach, innovative time-calibrated phyloanatomic tree construction, and a novel model of compartmentalized intra-host population dynamics provide a new perspective on the persistence of HIV-1 in the body. Achieving optimal cellular pharmacokinetics and spatial distribution of antiretroviral drugs in lymphoid tissue to fully suppress viral replication and preserve immune function would be a prerequisite to the elimination of the viral reservoir and ultimately a step towards a cure for HIV-1 infection.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 July; accepted 18 December 2015.

Published online 27 January 2016.

- Perelson, A. S. *et al.* Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**, 188–191 (1997).
- Chun, T. W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**, 183–188 (1997).
- Finzi, D. *et al.* Latent infection of CD4<sup>+</sup> T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nature Med.* **5**, 512–517 (1999).
- Finzi, D. *et al.* Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**, 1295–1300 (1997).
- Wong, J. K. *et al.* Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* **278**, 1291–1295 (1997).
- Brenchley, J. M. *et al.* T-cell subsets that harbor human immunodeficiency virus (HIV) *in vivo*: implications for HIV pathogenesis. *J. Virol.* **78**, 1160–1168 (2004).
- Chomont, N. *et al.* HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature Med.* **15**, 893–900 (2009).
- Zhu, T. *et al.* Evidence for human immunodeficiency virus type 1 replication *in vivo* in CD14<sup>+</sup> monocytes and its potential role as a source of virus in patients on highly active antiretroviral therapy. *J. Virol.* **76**, 707–716 (2002).
- Persaud, D. *et al.* Continued production of drug-sensitive human immunodeficiency virus type 1 in children on combination antiretroviral therapy who have undetectable viral loads. *J. Virol.* **78**, 968–979 (2004).
- Shen, L. & Siliciano, R. F. Viral reservoirs, residual viremia, and the potential of highly active antiretroviral therapy to eradicate HIV infection. *J. Allergy Clin. Immunol.* **122**, 22–28 (2008).
- Persaud, D. *et al.* A stable latent reservoir for HIV-1 in resting CD4<sup>+</sup>T lymphocytes in infected children. *J. Clin. Invest.* **105**, 995–1003 (2000).
- Yuki, S. A. *et al.* The distribution of HIV DNA and RNA in cell subsets differs in gut and blood of HIV-positive patients on ART: implications for viral persistence. *J. Infect. Dis.* **208**, 1212–1220 (2013).
- Fletcher, C. V. *et al.* Persistent HIV-1 replication is associated with lower antiretroviral drug concentrations in lymphatic tissues. *Proc. Natl Acad. Sci. USA* **111**, 2307–2312 (2014).
- Günthard, H. F. *et al.* Evolution of envelope sequences of human immunodeficiency virus type 1 in cellular reservoirs in the setting of potent antiretroviral therapy. *J. Virol.* **73**, 9404–9412 (1999).
- Kearney, M. F. *et al.* Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. *PLoS Pathog.* **10**, e1004010 (2014).
- Josefsson, L. *et al.* Majority of CD4<sup>+</sup> T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. *Proc. Natl Acad. Sci. USA* **108**, 11199–11204 (2011).
- Wong, J. K. *et al.* *In vivo* compartmentalization of human immunodeficiency virus: evidence from the examination of pol sequences from autopsy tissues. *J. Virol.* **71**, 2059–2071 (1997).
- von Andrian, U. H. & Mempel, T. R. Homing and cellular traffic in lymph nodes. *Nature Rev. Immunol.* **3**, 867–878 (2003).
- Archer, J. *et al.* Analysis of high-depth sequence data for studying viral diversity: a comparison of next generation sequencing platforms using Segminator II. *BMC Bioinformatics* **13**, 47 (2012).
- Kim, E. Y. *et al.* Human APOBEC3 induced mutation of human immunodeficiency virus type-1 contributes to adaptation and evolution in natural infection. *PLoS Pathog.* **10**, e1004281 (2014).
- Lemey, P., Rambaut, A. & Pybus, O. G. HIV evolutionary dynamics within and among hosts. *AIDS Rev.* **8**, 125–140 (2006).
- Frenkel, L. M. *et al.* Multiple viral genetic analyses detect low-level human immunodeficiency virus type 1 replication during effective highly active antiretroviral therapy. *J. Virol.* **77**, 5721–5730 (2003).
- Nickle, D. C. *et al.* Evolutionary indicators of human immunodeficiency virus type 1 reservoirs and compartments. *J. Virol.* **77**, 5540–5546 (2003).

24. Murrell, B. *et al.* Gene-wide identification of episodic selection. *Mol. Biol. Evol.* (2015).
25. Kryazhimskiy, S. & Plotkin, J. B. The population genetics of dN/dS. *PLoS Genet.* **4**, e1000304 (2008).
26. Mugal, C. F., Wolf, J. B. & Kaj, I. Why time matters: codon evolution and the temporal dynamics of dN/dS. *Mol. Biol. Evol.* **31**, 212–231 (2014).
27. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
28. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLOS Comput. Biol.* **5**, e1000520 (2009).
29. Vaughan, T. G., Kuhnert, D., Poppinga, A., Welch, D. & Drummond, A. J. Efficient Bayesian inference under the structured coalescent. *Bioinformatics* **30**, 2272–2279 (2014).
30. Frost, S. D. W. *et al.* Eight challenges in phylodynamic inference. *Epidemics* **10**, 88–92 (2015).
31. Tobin, N. H. *et al.* Evidence that low-level viremias during effective highly active antiretroviral therapy result from two processes: expression of archival virus and replication of virus. *J. Virol.* **79**, 9625–9634 (2005).
32. Anderson, J. A. *et al.* Clonal sequences recovered from plasma from patients with residual HIV-1 viremia and on intensified antiretroviral therapy are identical to replicating viral RNAs recovered from circulating resting CD4<sup>+</sup> T cells. *J. Virol.* **85**, 5220–5223 (2011).
33. Maldarelli, F. *et al.* HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179–183 (2014).
34. Wagner, T. A. *et al.* Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* (2014).
35. Althaus, C. L., Joos, B., Perelson, A. S. & Günthard, H. F. Quantifying the turnover of transcriptional subclasses of HIV-1-infected cells. *PLOS Comput. Biol.* **10**, e1003871 (2014).
36. Kepler, T. B. & Perelson, A. S. Drug concentration heterogeneity facilitates the evolution of drug resistance. *Proc. Natl Acad. Sci. USA* **95**, 11514–11519 (1998).
37. Rong, L., Dahari, H., Ribeiro, R. M. & Perelson, A. S. Rapid emergence of protease inhibitor resistance in hepatitis C virus. *Sci. Transl. Med.* **2**, 30ra32 (2010).
38. McLean, A. R. & Nowak, M. A. Competition between zidovudine-sensitive and zidovudine-resistant strains of HIV. *AIDS* **6**, 71–79 (1992).
39. Furtado, M. R. *et al.* Persistence of HIV-1 transcription in peripheral-blood mononuclear cells in patients receiving potent antiretroviral therapy. *N. Engl. J. Med.* **340**, 1614–1622 (1999).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank G. J. Beilman, A. Thorkelson, P. Swantek, K. Mars and K. Kunstman for their technical assistance. We thank E. Domingo and T. Bhattacharya for their constructive and informed review. We are indebted to the patients who participated in this study. This work was supported by the National Institutes of Health (DA033773 to S.M.W., AI1074340 to T.W.S., and GM110749 to S.L.K.P.), the Medical Research Council (G1000196 to M.H.M.), the Framework Programme for Research and Technological Development (278433-PREDEMICS to A.R.) and the European Research Council (260864 to A.R.). The Oxford Martin School supports H.R.F. All Souls College supports A.R.M. where S.M.W. held a Visiting Fellowship. A Newton International Fellowship from the Royal Society supported T.B. A Wellcome Trust Investigator award supported M.H.M. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Author Contributions** T.W.S., A.T.H., and S.M.W. conceived the experiments and designed the study. T.W.S. and J.G.C. acquired the patient tissue samples. A.T.H. performed the *in situ* hybridization experiments. C.V.F. measured the intracellular drug concentrations. R.L.-R., E.-Y.K., and Y.-S.C. generated the viral sequences. R.L.-R., T.B., E.-Y.K., S.P., M.H.M., S.L.K.P., A.R., and S.M.W. analysed the data. R.L.-R., T.B., J.A., and A.R. conducted the Bayesian inference analyses. H.R.F., A.R.M., and S.M.W. developed the spatial and dynamic model. R.L.-R., T.B., H.R.F., A.T.H., S.L.K.P., A.R.M., A.R., and S.M.W. wrote the paper, with extensive input from all authors. All authors discussed the results and commented on the manuscript.

**Author Information** Nucleotide sequence alignments were deposited in GenBank with the accession numbers KT829617–KT831260. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.W. ([s-wolinsky@northwestern.edu](mailto:s-wolinsky@northwestern.edu)).

## METHODS

**Data reporting.** No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during outcome assessment.

**Study subjects.** The three study subjects were enrolled into a clinical protocol where treatment was started after we obtained peripheral blood by venepuncture, inguinal lymph nodes by excisional biopsy, and ileum and rectal biopsies through colonoscopy<sup>13</sup>. Two subjects were antiretroviral drug naive (1727 and 1679) and one subject (1774) had not received any drugs for at least 1 year before enrolment. Subjects 1679 and 1774 received emtricitabine (FTC), tenofovir (TFV; as the disoproxil fumarate (TDF)), and atazanavir with ritonavir (ATV/R). Subject 1727 received FTC, TDF and efavirenz (EFV). In all three subjects, conventional typing methods confirmed that the plasma virus was sensitive to the potent antiretroviral regimen that they received. Subjects 1727 and 1679 were well-suppressed patients. Subject 1774 continued to have measureable amounts of HIV-1 RNA in plasma after 3, but not 6 months of treatment. We obtained peripheral blood, inguinal lymph node, and terminal ileum and rectum biopsies once again at 3 and 6 months after starting treatment. Laboratory procedures for tissue management, *in situ* hybridization, and analytical pharmacology are described elsewhere<sup>13</sup>.

The Institutional Review Board of the University of Minnesota approved the study. All subjects provided written informed consent.

**Extraction and quantification of viral nucleic acids.** Nucleic acids were extracted from frozen cells obtained from blood or lymphoid tissue using the MasterPure Complete DNA and RNA Purification Kit (Epicentre). Viral RNA was isolated from plasma using the PureLink Viral RNA/DNA Mini Kit (Life Technologies). HIV-1 was quantified using a quantitative reverse transcription PCR assay. The relative amount of HIV-1 target-DNA was normalized to the quantification cycle for a concentration calibrator by using an external standard curve of serial tenfold dilutions of reference DNA for the Gag region of HIV-1 derived from the plasmid pNL-43. All reactions were performed in triplicate on the ABI 7900HT sequence detector (Applied Biosystems).

**Library preparation for deep sequencing.** As the number of viral templates in the sampled material is low, we used an amplicon-based deep sequencing strategy. To minimize biased amplification of the target sequence, primer locations in the Gag and Pol regions of HIV-1 were selected on the basis of the alignment positional entropy in the multiple sequences aligned from the Los Alamos National Laboratory HIV-1 sequence database (<http://www.hiv.lanl.gov/>). The primers were computationally screened for cross-dimer interactions and the concentration of each primer was optimized for amplification. For each sample, blanks were included to screen for contamination. We used designated, physically separated areas within the laboratory to set-up PCR, which avoided contact with potentially contaminating amplicons.

To generate the long read-length PCR amplicons sequenced in this study (range, 509–587 bp read-lengths per run depending on the gene region being analysed), we amplified the Gag and Pol regions of HIV-1. For the Gag region of HIV-1, we used forward primer gag\_632F\_EK (5'-GCAGTGGCGCCCGAAC-3' (corresponding to HXB2 nucleic acid sequence numbering positions 632–647) and reverse primer gag\_1788R\_EK (5'-AATAGTCTTACAATCTGGGTTCGC-3' (1788–1765)). For the Pol region of HIV-1 that spanned the genomic region encoding the viral enzyme protease and reverse transcriptase, we used forward primer HIV-Pro1\_2137F (5'-CAGAGCAGACCAGAGCCAAC-3', corresponding to positions 2137–2156) and reverse primer HIV-RT1\_3531R (5'-CTGCTATTAGTCTTTTGATGGGTC-3' (3531–3507)). PCR amplification was performed using the High Fidelity Platinum Taq DNA Polymerase (Invitrogen) with thermal cycling conditions of 94 °C for 2 min, followed by 35 cycles of 94 °C for 15 s, 54 °C for 15 s, 68 °C for 1 min, with a final extension step at 68 °C for 5 min.

We used an integrated sequencing pipeline for library construction, template amplification, and DNA sequencing as in ref. 20. Multiplex Identifiers were included during library preparation for sample barcoding. For the Gag region of HIV-1, we used the forward primer A-Gag\_977F\_degEK 5'-primer A-GCTACAACCAKCCCTYACAGAC-3' (977–1000) and the reverse primer B-Gag\_1564R\_degEK 5'-primer B-CTACTGGGATAGGTGGATTAYKTG-3' (1564–1541) to generate a 587 bp amplicon (977–1564). For the Pol region of HIV-1 that spanned the genomic region encoding the viral enzyme protease, we used forward primer A-Pol1\_2235F 5'-primer-A-ACTGTATCCTTTAGCT TCCCTCA-3' (2235–2262) and the reverse primer B-Pol1\_2744R 5'-primer B-TTTCTTTATGGCAAATACTGGAG-3' (2744–2721) to generate a 509 bp amplicon (2235–2744). For the Pol region of HIV-1 that spanned the genomic region encoding the viral enzyme reverse transcriptase, we used forward primer A-Pol2\_2700F 5'-primer-A-GGGCCTGAAAATCCATACAAT-3' (2700–2721) and the reverse primer B-Pol2\_3265R 5'-primer-B-CATTATACAGGATGGAGT TCATA-3' (3265–3242) to generate a 565 bp amplicon (2700–3265). PCR was performed as detailed above. DNA amplicon libraries were resolved on a pre-cast

2% agarose gel and purified with QIAquick Gel Extraction kit (Qiagen) and AMPure XP SPRI beads (Beckman Coulter).

Libraries were quantified with KAPA Library Quant Kit (Kapa Biosystems) on Agilent 2100 Bioanalyzer High Sensitivity DNA chip (Agilent) for concentration and size distribution. The concentration of the product DNA was normalized before pooling to achieve sequence uniformity across amplicons. Controls were spiked into the reaction to monitor the library construction process and potential index cross-contamination. The known internal control sequence (clonal sequence of 456 bp) introduced into the reaction was used to calculate the single nucleotide error rate and set the cut-off for the sequence analysis where no control errors could be detected. PCR errors are more common in later cycles of amplification, and being limited to small copy numbers they have little impact on the haplotype distribution. Bidirectional sequencing using the 454 Life Sciences' GS-FLX sequencing platform (Roche) provided independent confirmation of sequence information. The long read-length sequences were sorted based on index sequences, trimmed to remove residual adaptor bases from the ends of the reads, and filtered for length and duplicates before alignment. Read depth and coverage estimation that met predetermined coverage thresholds were performed as in ref. 20. Sequencing quality metrics were calculated for all samples using FastQC and only high-quality sequencing libraries were used in the ensuing analyses.

**Sequence clean up and assembly.** Experimental precision along with deep coverage allows for accurate estimation of the underpinning diversity of the virus population. We binned the sequence reads by multiplex identifier barcodes, and identified and excluded sequencing errors and misaligned regions from the analysis by computational methods. A small number of reads had a disproportionate number of errors that accounted for most of the inaccuracy in the full data set. After quality filtering and trimming to a uniform length, we proceeded to build the different haplotypes present in each of the samples. We began by collapsing the identical sequences into haplotypes using reference-guided assembly to avoid the use of uninformative sequence repeats. Haplotypes at prevalence above the error threshold (defined by the internal control spiked in the sequencing runs), which corresponded to variants present above 0.04% of the total existing variants in the collapsed alignments, were used in the analysis.

For the processing of the temporally and spatially linked deep-sequencing data for studying viral diversity, the viral sequences were first aligned against the HXB2 reference sequence (GenBank accession number K03455) using Segminator II (version 0.1.1). We then generated a consensus viral sequence for each patient as a reference for assembly to improve the alignment quality. A statistical model that utilizes platform error rates in conjunction with patterns within nucleotide frequencies derived from data obtained from related samples (that is, different compartments within the host or temporally linked samples) was used to separate low frequency platform error from true variation. This model is termed 'probabilistic read error detection across temporally obtained reads' (PREDATOR) and is implemented within Segminator II. This statistical framework maintains the reading frame and corrects for deep sequencing errors<sup>19</sup>. We corroborated the number of haplotypes and the frequency of haplotypes that explain the data using a second reconstruction algorithm based on combinations of multinomial distributions to analyse the k-mer frequency spectrum of the sequencing data implemented with QuRe<sup>40</sup>. We screened the sequence alignments for recombinant sequences using the GARD algorithm implemented in HyPhy<sup>41,42</sup>.

**Sampling variance.** The high coverage of massively parallel sequencing is necessary to ensure reliable detection of low-frequency viral variants. A simple calculation based on the geometric distribution shows that in order to guarantee that a viral template occurring at frequency  $f$  is detected with probability  $p$  or better, it is necessary to sequence at least  $\log(1-p)/\log(1-f) - 1$  templates. A variant of frequency 0.01 (that is, 1%), for example, would require 450 sequences to ensure its detection at probability 0.99 (that is, 99%) or better. Conversely, a study using 100 single-genome sequences would detect a variant of frequency 0.01 with probability 0.64. A low number of input DNA templates derived from one compartment that catches the spillover from another does not account for the complexities of partial observation and spatial heterogeneity that could lead to measurement error elsewhere<sup>16</sup>. This complication emphasizes the challenge in trying to extrapolate from single-template sequencing the magnitude and character of the virus population that comprises the viral reservoir.

**Maximum-likelihood tree construction.** Maximum-likelihood phylogenies were created with PhyML using the general time-reversible model with the proportion of invariant sites and gamma distribution of among-site rate variation (GTR + I +  $\Gamma_4$ ) nucleotide substitution model<sup>43</sup> applying an approximate likelihood ratio test for branch support<sup>44</sup>. We estimated trees on viral sequence sets from which gaps in the alignment were removed and considered as missing data for the reason that maximum-likelihood tree error may increase with inclusion of unreliable sites. We assessed the temporal structure of the trees by performing linear regression on the root-to-tip distances of samples versus the time of sampling and tested the validity



of the time-dependency of the evolution rate estimates with the assumption of a strict molecular clock using the program Path-O-Gen v1.4 (<http://tree.bio.ed.ac.uk/software/pathogen/>). We used the Highlighter sequence visualization tool (<http://www.HIV.lanl.gov>) to trace commonality between sequences in an alignment based on individual nucleotide changes.

**Compartmentalization.** We tested for subdivision of viral sequences into sub-populations in the different compartments at each time point. We calculated genetic distances using the Wright's  $F_{ST}$  and  $S_{im}$  test statistics<sup>45,46</sup>. We used a bootstrap test to determine the confidence of the estimates and performed a permutation test (1000 iterations) to assess the significance levels of the obtained scores.

**Identifying selection.** We used a modification of a random effects branch-site model to detect positive selection and test whether the phylogeny diverged over time<sup>47</sup>. A likelihood ratio hypothesis test compared the fit of the model using a 3-bin  $\omega$  distribution ( $\omega_1$  and  $\omega_2$  in  $[0,1]$ ,  $\omega_3$  unrestricted) to describe the evolution of all branches in the tree, to the fit of the model where  $\omega_3$  is also restricted to be in  $[0,1]$ . We tested whether or not a proportion of sites along internal branches of the intra-host viral phylogeny have been subject to episodic selection ( $\omega > 1$ )<sup>24</sup>, restricting the test to internal branches to lessen the biasing effects of neutral or deleterious mutations on  $\omega$  estimates<sup>48,49</sup> and serve as a proxy for population level selection<sup>25,26</sup>.

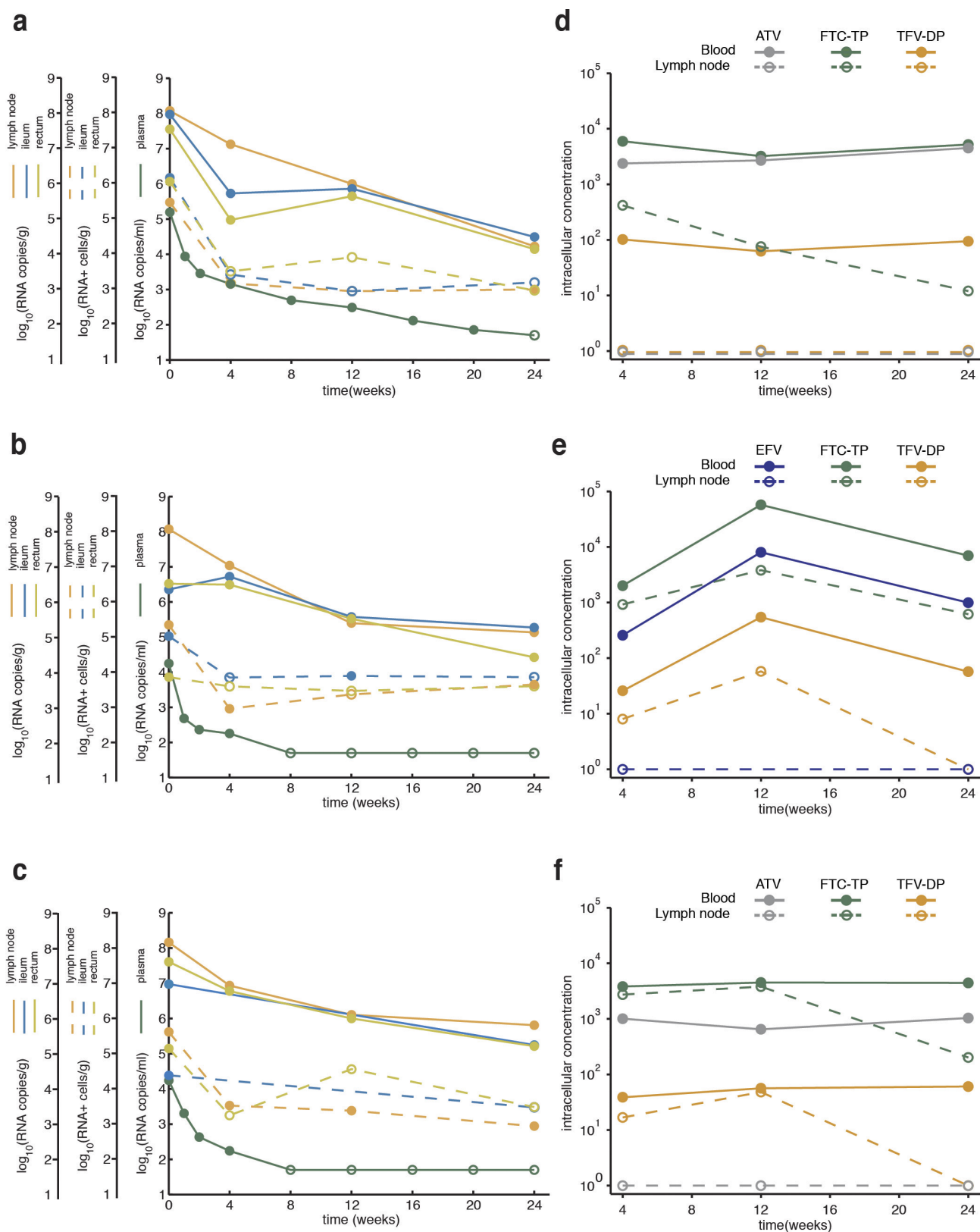
**Time-calibrated phylogenetic tree construction.** To resolve the phyloanatomy, we reconstructed the temporal and spatial dynamics of the viral haplotype lineages with a Bayesian statistical framework using Markov chain Monte Carlo sampling for evolutionary hypothesis testing, as implemented in BEAST version 2.1.2 (ref. 27). This approach was used to sample phylogenies from their joint posterior distribution, in which the viral haplotypes are restricted by their known date of sampling, using a simple substitution model described by Hasegawa–Kishino–Yano (HKY) to avoid over-parameterization<sup>50</sup>. Models differing in assumptions on mutation rate and effective population size were run for 100 million generations each and compared using the Bayes factor as implemented in Tracer version 1.6. We determined that the best-fit model included a strict molecular clock and assumed a constant population size. We used a symmetric transition model with constant rates over time that considered a discretized diffusion process among the different compartments. This was formalized as a continuous time Markov chain model to reconstruct the spatial dynamics between compartments. All chains were run for sufficient length and convergence of the relevant parameters was assessed using Tracer version 1.6, ignoring 10% of the chain as burn-in.

We summarized the connections between virus evolution and anatomical compartment history using an annotated MCC phylogenetic tree estimated with

BEAST. The model and its parameters were chosen after computing the posterior probability of several models to obtain the discriminatory Bayes factors. Because population structure, whether due to spatial segregation or limitations to gene flow, may affect evolutionary dynamics, we confirmed that the direction of flow was not due to oversampling of a particular environment, by running a two-deme Bayesian inference under a structured coalescent model with a HKY substitution model assuming a strict molecular clock<sup>29</sup>, which is less susceptible to sampling issues than our trait-based analysis. For completeness, we conducted a search for topologies and divergence times assuming a relaxed molecular clock as well. In the analyses performed, HIV-1 showed a high degree of clock-like evolution and a mean nucleotide substitution rate expected to be within the bounds necessary to obtain meaningful phyloanatomic information from sequence data. Using the location of each of the haplotypes, a discrete trait was included in the inference. We used BEAST to estimate the probabilities of each of the possible states.

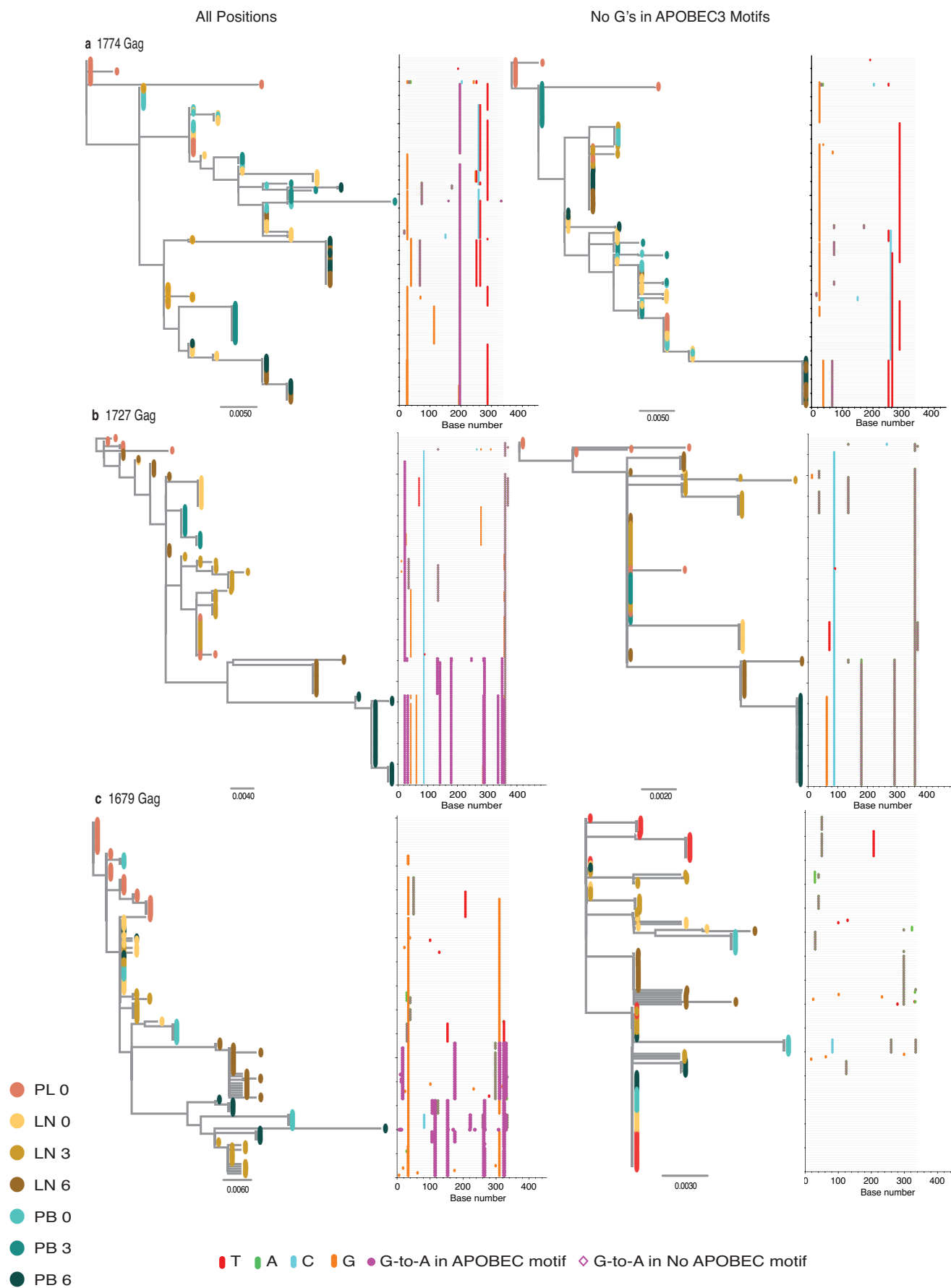
**Statistical analysis.** Standard descriptive statistics were performed with the use of the STATA, GraphPad, or R packages.

40. Prosperi, M. C. & Salemi, M. QuRe: software for viral quasispecies reconstruction from next-generation sequencing data. *Bioinformatics* **28**, 132–133 (2012).
41. Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
42. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* **23**, 1891–1901 (2006).
43. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
44. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552 (2006).
45. Hudson, R. R., Slatkin, M. & Maddison, W. P. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589 (1992).
46. Hudson, R. R. A new statistic for detecting genetic differentiation. *Genetics* **155**, 2011–2014 (2000).
47. Kosakovsky Pond, S. L. *et al.* A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* **28**, 3033–3043 (2011).
48. Kosakovsky Pond, S. L. *et al.* Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLOS Comput. Biol.* **2**, e62 (2006).
49. Pybus, O. G. *et al.* Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol. Biol. Evol.* **24**, 845–852 (2007).
50. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).



**Extended Data Figure 1 | The amount of virus and concentration of drugs measured during antiretroviral therapy. a–c,** The changes in copies of HIV-1 RNA per ml of blood, the number of the HIV-1 RNA particles bound to the follicular dendritic cell network per gram of lymphoid tissue, and the number HIV-1 RNA positive cells per gram of lymphoid tissue over the first 6 months of treatment in subjects 1774, 1727 and 1679 (a, b, and c, respectively). Filled circles represent detectable measures. Unfilled circles represent undetectable measures and are plotted at the limit of detection. **d–f,** Antiretroviral drug concentrations

in cells from lymph node (dashed line) or blood (solid line) in subjects 1774, 1727 and 1679 (d, e and f, respectively; see Methods). Intracellular TFV-diphosphate (TFV-DP) concentrations (fmol per 10<sup>6</sup> cells) are shown in orange, FTC-triphosphate (FTC-TP) (fmol per 10<sup>6</sup> cells) in green, ATV (ng ml<sup>-1</sup>) in purple, and EFV (ng ml<sup>-1</sup>) in blue. Samples with concentrations that were below the limits of quantification (2.5 fmol per 10<sup>6</sup> cells, 2.5 fmol per 10<sup>6</sup> cells, 0.014 ng ml<sup>-1</sup> and 0.063 ng ml<sup>-1</sup>, respectively) were assigned a value of 1 for graphical illustration purposes.

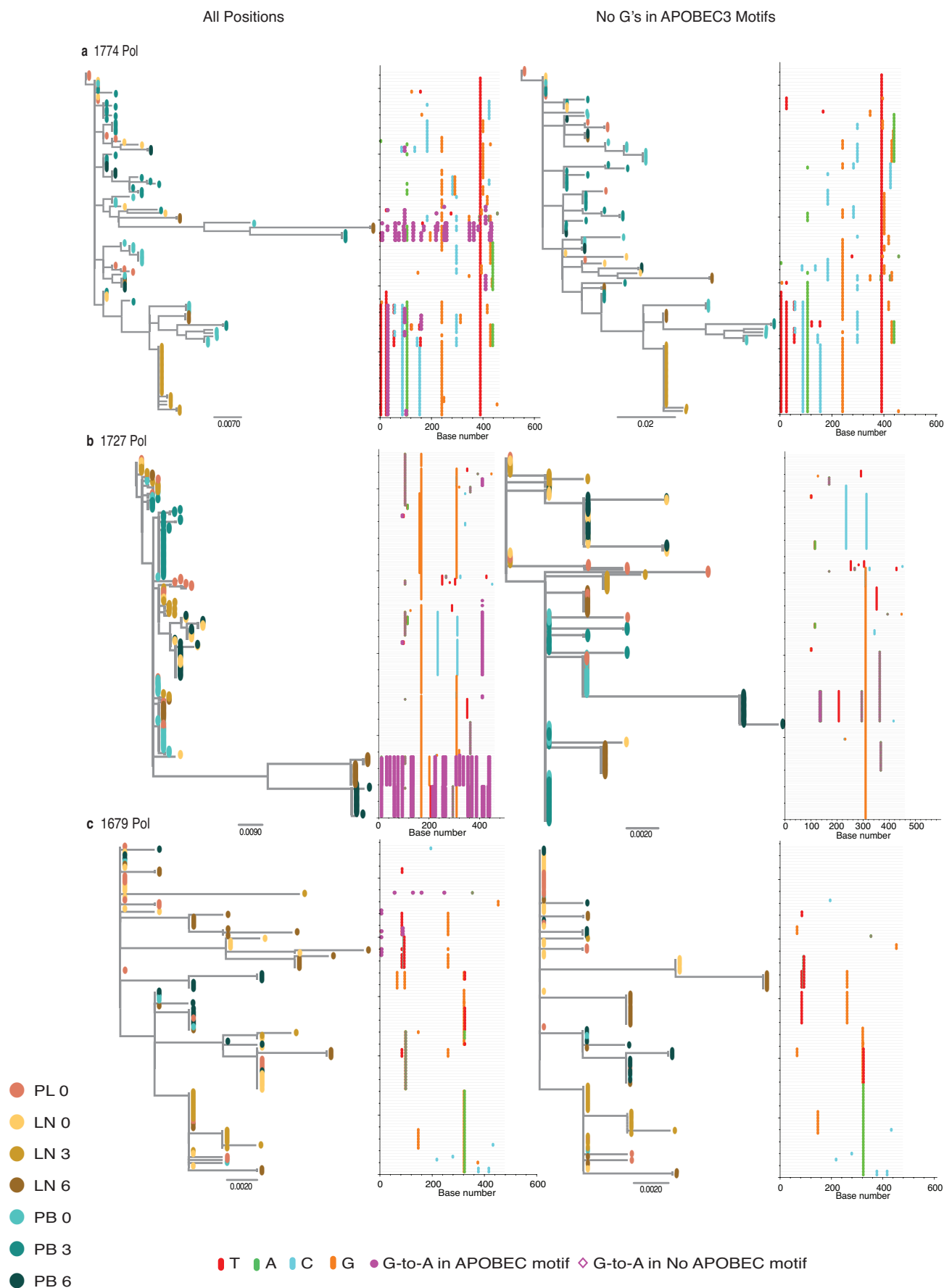


Extended Data Figure 2 | See next page for caption.



**Extended Data Figure 2 | Phylogenies and Highlighter plots for the Gag region of HIV-1. a–c,** Maximum-likelihood trees were constructed using gene sequences from the Gag region of HIV-1 from lymph node and blood before and after the guanosines within all possible APOBEC3 trinucleotide sequence context of edited sites were masked in the alignments, regardless of their presence in hypermutated or non-hypermutated sequences, to avoid their distortion in the phylogenetic reconstructions. Branch tips are coloured according to compartment sampled: red for plasma, gold for lymph node, and blue for blood. The progressive shading of the colours of the branch tips indicate the points in time sampled. Phylogenetic trees reconstructed from the haplotypes in which the guanosines in the APOBEC3 trinucleotide context of the edited sites are masked in the alignments correct the skewing effect caused by clustering of shared haplotypes that harbour repetitive G-to-A substitutions and longer branch lengths caused by a larger number of these mutations in the hypermutated sequences while retaining the phylogenetic information. The horizontal

scale indicates the expected number of substitutions per nucleotide site per unit time, with haplotypes from later time points having diverged more. The Highlighter plots show the haplotypes from the lymphoid tissue and blood time point clusters aligned to the plasma virus sequence from day 0. The particular nucleotide changes are colour-coded in the alignment (thymidine, red; adenosine, green; cytosine, blue; and guanosine, orange). Magenta circles represent APOBEC3-induced G-to-A change in a trinucleotide context of the edited sites, which are distinguishable from the more random error-prone viral reverse transcriptase and RNA polymerase II replicating enzyme induced mutations<sup>6</sup>. Gene sequences from the Gag region of HIV-1 from subject 1774, who continued to have measureable amounts of HIV-1 RNA in plasma on treatment, and subjects 1727 and 1679 who were well-suppressed on treatment (**a**, **b** and **c**, respectively) before and after the guanosines within the particular APOBEC3 trinucleotide sequence context of edited sites were masked in the entire sequence alignment (left and right panels, respectively).

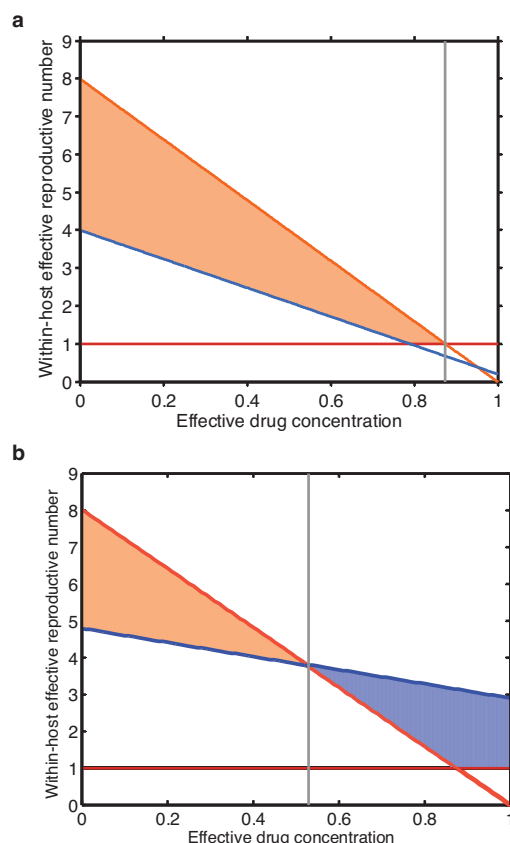


Extended Data Figure 3 | See next page for caption.

**Extended Data Figure 3 | Phylogenies and Highlighter plots for the Pol region of HIV-1.** **a–c**, Maximum-likelihood trees were constructed using gene sequences from the Pol region (reverse transcriptase (Pol2)) of HIV-1 from lymph node and blood before and after the guanosines within all possible APOBEC3 trinucleotide sequence context of edited sites were masked in the alignments, regardless of their presence in hypermutated or non-hypermutated sequences, to avoid their distortion in the phylogenetic reconstructions. Branch tips are coloured according to compartment sampled: red for plasma; gold for lymph node; and blue for blood. The progressive shading of the colours of the branch tips indicate the points in time sampled. The horizontal scale indicates the expected number of substitutions per nucleotide site per unit time with haplotypes

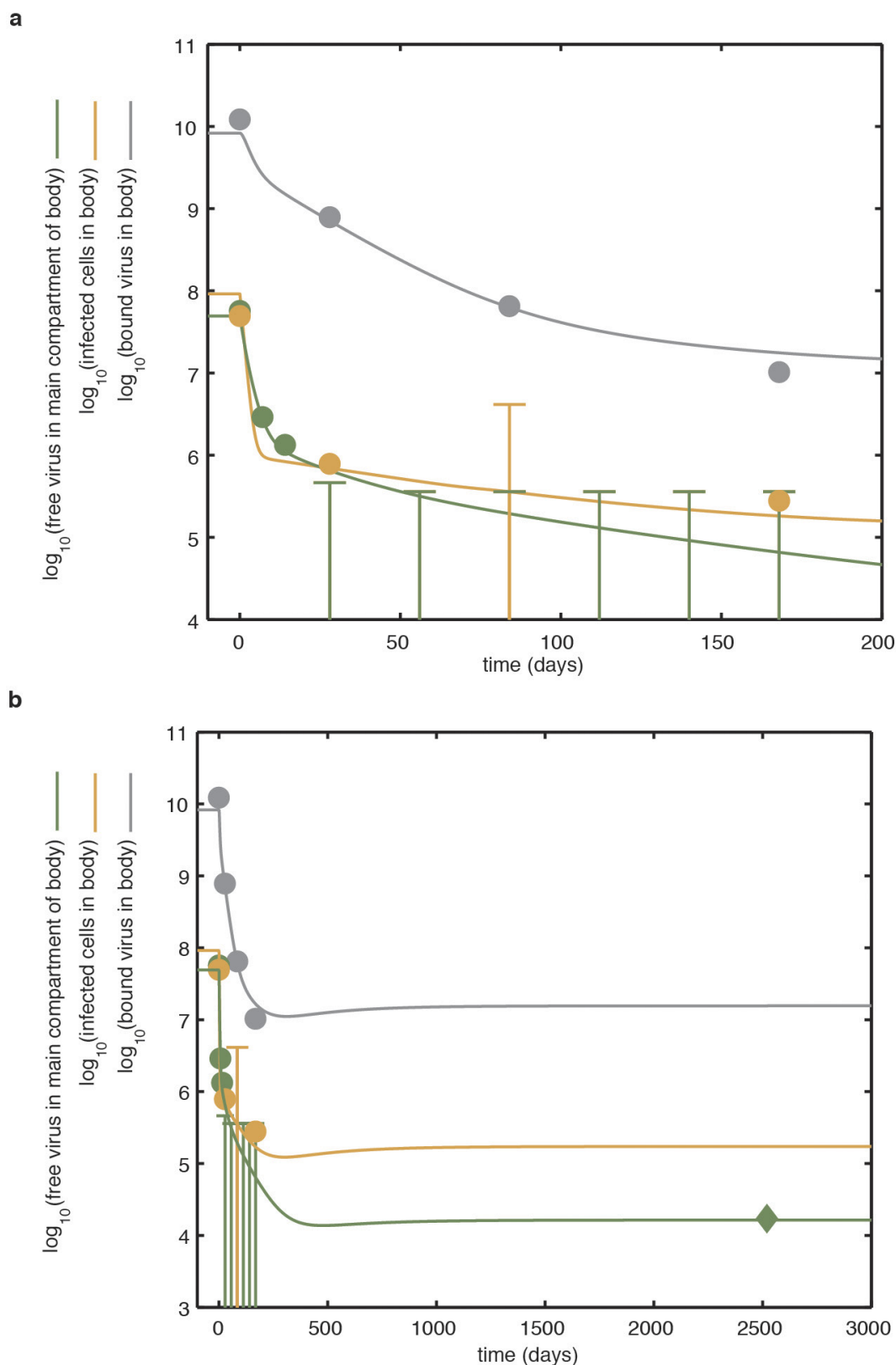
from later time points having diverged more. The Highlighter plots show the haplotypes from the lymphoid tissue and blood time point clusters aligned to the plasma virus sequence from day 0. The particular nucleotide changes are colour-coded in the alignment (thymidine, red; adenosine, green; cytosine, blue; and guanosine, orange). Magenta circles represent APOBEC3-induced G-to-A change in a trinucleotide context of the edited sites. Gene sequences from the Pol region of HIV-1 that spanned the genomic region encoding the viral enzyme reverse transcriptase from subjects 1774, 1727, and 1679 (**a**, **b** and **c**, respectively) before and after the guanosines within the particular APOBEC3 trinucleotide sequence context of edited sites were masked in the entire sequence alignment (left and right panels, respectively).





#### Extended Data Figure 4 | Alternative drug-dependent fitness landscape plots.

**a.** Fitness landscape plot for a partially drug-resistant strain. This strain confers a low level of drug resistance relative to the replicative fitness cost imposed by the resistance mutations. The drug-resistant strain (blue line) does not out-compete the drug-sensitive strain (orange line) at any effective treatment concentration where it can grow. There are two phases to the dynamics: at lower effective drug concentrations (left of grey line), the drug-sensitive strain thrives; beyond this threshold, neither strain can continuously replicate. **b.** Fitness landscape plot for a highly drug-resistant strain. This strain confers a high-level of drug resistance relative to the replicative fitness cost imposed by the resistance mutations. At low effective drug concentrations (left of grey line), the drug-sensitive strain out-competes the drug-resistant strain. At high effective drug concentrations, the drug-resistant strain out-competes the drug-sensitive strain and can continuously replicate. We argue that, typically, highly drug-resistant mutants of this sort neither exist in the viral population of patients before treatment, nor arise through random mutation during the course of antiretroviral therapy (see Supplementary Information and Supplementary Table 2). Drug-resistant strains that are capable of ongoing replication at high effective drug concentrations are not typically generated in individuals because: they are generated in a single step very rarely; and stepwise generation from partially resistant strains is also rare because partially resistant strains are out-competed in the sanctuary site that constantly replenishes the pool. The strain-specific effective reproductive numbers for the drug-sensitive (orange line) and drug-resistant (blue line) strains are shown. The orange sections highlight regions of parameter space where the drug-sensitive strain will dominate, and the blue section highlights regions of parameter space where the drug-resistant strain will dominate. For simplicity, only the impact of changes to the effectiveness of a single drug in a single compartment is shown.



**Extended Data Figure 5 | Model of replication dynamics and treatment effectiveness in the viral reservoir fitted to the data.** The model is fitted to the total inferred average body counts of free virus particles (green line), infected CD4<sup>+</sup> T cells (orange line) and virus bound to the follicular dendritic cell network of B-cell follicles (grey line). **a**, Dynamics over the first 200 days of treatment. Note that early in antiretroviral therapy, HIV-1 RNA in plasma declines more rapidly than virus bound to the follicular dendritic cell network of B-cell follicles. Circles demonstrate

average data from the three patients discussed in detail in this study and an additional nine patients presented elsewhere<sup>13</sup>. Where the average value was indeterminate because of test sensitivity, the data are fitted below the upper limit of the average log<sub>10</sub> infectious units. The range below the upper limit is represented by a vertical bar. **b**, Dynamics over a longer period. The model predicts the persistent low-level viral RNA in plasma. (see Supplementary Information). The optimal model fit parameters are presented in Supplementary Table 1.

**Extended Data Table 1 | The genetic distance measured between the haplotypes in the Gag or Pol regions of HIV-1**

Subject	Compartment	Region	time = 0	time = 6 months	P-value
1774	LN	Gag	4.96E-03	1.14E-02	<0.0001
		Pol1	6.38E-03	5.58E-03	0.3783
		Pol2	5.85E-03	1.26E-02	0.0208
	PB	Gag	4.20E-03	1.20E-02	<0.0001
		Pol1	1.83E-03	1.20E-02	<0.0001
		Pol2	9.72E-03	1.07E-02	0.0070
1727	LN	Gag	5.73E-05	8.59E-03	<0.0001
		Pol1	1.35E-03	1.59E-02	<0.0001
		Pol2	1.08E-03	5.26E-03	<0.0001
	PB	Gag	NA	NA	NA
		Pol1	1.87E-03	5.26E-03	0.3280
		Pol2	1.16E-03	1.01E-02	<0.0001
1679	LN	Gag	6.68E-04	3.85E-04	0.5050
		Pol1	1.73E-04	2.16E-03	<0.0001
		Pol2	1.54E-03	3.28E-03	0.0023
	PB	Gag	4.35E-03	5.97E-03	0.8799
		Pol1	2.23E-03	4.36E-03	0.0989
		Pol2	0.00E+00	1.57E-03	0.2383

We used sequence data from the Gag or Pol regions (protease (Pol1) and reverse transcriptase (Pol2)) of HIV-1 (range, 315–487 bp) to perform a genetic distance analysis. We masked the guanosines in APOBEC3 trinucleotide contexts of the edited sites from the alignments. We compared the proportion of substitutions per site (*p*-distance) for haplotypes within lymph node (LN) or peripheral blood (PB) from each subject at 6 months to the most common haplotype present at day 0. Most comparisons were statistically significant (Mann–Whitney *U* test, *P*-values <0.1), indicating ongoing diversification of the viral populations.



**Extended Data Table 2 | HIV-1 evolutionary rate calculated after removing hypermutated haplotypes**

Subject	Region	Compartment	$\mu$	Std Err
1774	Gag	LN	1.54E-03	1.42E-05
		PB	1.55E-03	9.33E-06
	Pol1	LN	-4.87E-04	2.17E-05
		PB	1.58E-03	1.04E-05
	Pol2	LN	1.79E-03	2.44E-05
		PB	1.60E-04	3.90E-05
1727	Gag	LN	1.87E-03	1.49E-05
		PB	NA	NA
	Pol1	LN	1.81E-03	1.25E-05
		PB	6.19E-04	1.13E-05
	Pol2	LN	1.44E-03	1.36E-05
		PB	1.70E-03	6.56E-06
1679	Gag	LN	4.74E-04	2.25E-05
		PB	8.34E-04	5.09E-05
	Pol1	LN	1.05E-03	8.89E-06
		PB	3.90E-04	2.77E-05
	Pol2	LN	4.89E-04	2.20E-05
		PB	5.12E-04	7.56E-05

We used a linear regression model to estimate the evolutionary rate for the Gag or Pol regions (protease (Pol1) and reverse transcriptase (Pol2)) of HIV-1 within lymph node or peripheral blood of each subject. We calculated the slope ( $\mu$ ) of the linear regression between time and the direct pairwise genetic distances (number of substitutions per site per month) from the most common haplotype at day 0 for each subject. Haplotypes found to harbour G-to-A hypermutated sequences were removed from the analysis to limit the effect of inactivating mutations on the estimates. There is very strong evidence, now presented in numerous studies<sup>17</sup>, that acute HIV-1 infections are largely founded by a single (or at best a few) viral strains; this initial bottleneck is followed by rapid population diversification in the absence of treatment. The structures of our intra-host phylogenies at day 0 support this pattern. Even if multiple populations are transmitted and are able to establish infection and co-circulate, Bayesian phylodynamics models can properly account for this and estimate evolutionary rates accurately.

Extended Data Table 3 | Patterns of genetic divergence measured between viral haplotypes in lymphoid tissue and blood

**a**

Subject	Region	Day 0		Month 6	
		$S_{nn}$	Probability (Random $S_{nn}$ > Observed $S_{nn}$ )	$S_{nn}$	Probability (Random $S_{nn}$ > Observed $S_{nn}$ )
1774	Gag	0.71	0	0.53	0.086
	Pol1	0.71	0	0.91	0
	Pol2	0.86	0.001	0.90	0.003
1727	Gag	NA	NA	1.00	0
	Pol1	0.91	0	0.98	0
	Pol2	0.98	0	1.00	0
1679	Gag	0.72	0	0.97	0
	Pol1	0.75	0.04	0.77	0.011
	Pol2	0.86	0.003	0.86	0

**b**

Subject	Region	Day 0		Month 6	
		$F_{st}$	Probability (Random $F_{ST}$ > Observed $F_{ST}$ )	$F_{st}$	Probability (Random $F_{ST}$ > Observed $F_{ST}$ )
1774	Gag	0.13	0	-0.01	0.51
	Pol1	0.13	0	0.41	0
	Pol2	0.05	0.236	0.36	0
1727	Gag	NA	NA	0.76	0
	Pol1	0.51	0	0.58	0
	Pol2	0.73	0	0.34	0
1679	Gag	-0.06	0.75	0.67	0
	Pol1	0.37	0.018	0.02	0.394
	Pol2	0.23	0.154	0.35	0

$S_{nn}$  and  $F_{ST}$  pairwise genetic differentiation (**a** and **b**, respectively) between peripheral blood and lymph node samples for the Gag and Pol regions of HIV-1 from subjects 1774, 1727, and 1679 at day 0 and after 6 months of antiretroviral therapy.

Extended Data Table 4 | Episodic selection estimated across sites along internal branches of the phylogeny

Subject	Region	Branch-site $\omega$ distribution (internal branches)	Branch-site $\omega$ distribution (terminal branches)	P-value for episodic diversifying selection on internal branches
1774	Gag	$\omega_1 = 0.09$ (81%) $\omega_2 = 0.35$ (5%) $\omega_3 = 1.00$ (14%)	$\omega_1 = 0.09$ (69%) $\omega_2 = 0.15$ (14%) $\omega_3 = 0.32$ (17%)	1.0
	Pol1	$\omega_1 = 0.03$ (95%) $\omega_2 = 0.07$ (3%) $\omega_3 = 1.00$ (2%)	$\omega_1 = 0.09$ (99.9%) $\omega_2 = \infty$ (0.1%)	1.0
	Pol2	$\omega_1 = 0.11$ (89%) $\omega_2 = 0.35$ (11%)	$\omega_1 = 0.16$ (71%) $\omega_2 = 0.22$ (29%)	1.0
1727	Gag	$\omega_1 = 0.16$ (91%) $\omega_2 = 0.36$ (9%)	$\omega_1 = 0.00$ (87%) $\omega_2 = 5.00$ (13%)	1.0
	Pol1	$\omega_1 = 0.05$ (94%) $\omega_2 = 0.24$ (2%) $\omega_3 = 1.00$ (4%)	$\omega_1 = 0.09$ (60%) $\omega_2 = 0.21$ (13%) $\omega_3 = 0.25$ (27%)	1.0
	Pol2	$\omega_1 = 0.09$ (87%) $\omega_2 = 0.51$ (2%) $\omega_3 = 1.00$ (11%)	$\omega_1 = 0.13$ (99.6%) $\omega_2 = \infty$ (0.4%)	1.0
1679	Gag	$\omega_1 = 0.15$ (63%) $\omega_2 = 1.19$ (37%)	$\omega_1 = 0.68$ (100%)	1.0
	Pol1	$\omega_1 = 0.14$ (82%) $\omega_2 = 0.54$ (0.5%) $\omega_3 = 1.00$ (17%)	$\omega_1 = 0.19$ (97%) $\omega_2 = 0.15$ (2.8%) $\omega_3 = \infty$ (0.4%)	1.0
	Pol2	$\omega_1 = 0.19$ (94.8%) $\omega_2 = 0.93$ (4.9%) $\omega_3 = \infty$ (0.3%)	$\omega_1 = 0.31$ (100%)	$10^{-6}$

A branch-site evolutionary model provided a statistical framework to search for evidence of episodic selection on internal branches in the tree. The  $\omega$  distribution is given for both internal and external branches for the Gag or Pol regions (protease (Pol1) and reverse transcriptase (Pol2)) from each subject.



Extended Data Table 5 | Estimated migration rates between lymph nodes and blood using Bayesian inference under the structured coalescence

Subject	Migration rate	Mean	SEM
1774	From PB to LN	0.48	1.08E-02
	From LN to PB	1.70	1.57E-01
1727	From PB to LN	0.31	6.99E-03
	From LN to PB	0.57	1.21E-02
1679	From PB to LN	0.18	2.20E-03
	From LN to PB	0.71	1.06E-02

Migration rates (in fraction of emigrants per month) were estimated in the Gag region of HIV-1 for the three study subjects using Bayesian inference under the structured coalescence model, assuming a constant population size and a strict molecular clock. Migration rates from lymph node to blood and vice versa are shown with their standard error of the mean (SEM) after running at least 50 million Markov chain Monte Carlo sampling steps and reaching high values of estimated sample sizes for all parameters.

# Active medulloblastoma enhancers reveal subgroup-specific cellular origins

Charles Y. Lin<sup>1†\*</sup>, Serap Erkek<sup>2,3\*</sup>, Yiai Tong<sup>4</sup>, Linlin Yin<sup>5</sup>, Alexander J. Federation<sup>1</sup>, Marc Zapatka<sup>6</sup>, Parthiv Haldipur<sup>7</sup>, Daisuke Kawauchi<sup>3</sup>, Thomas Risch<sup>8</sup>, Hans-Jörg Warnatz<sup>8</sup>, Barbara C. Worst<sup>3</sup>, Bensheng Ju<sup>9</sup>, Brent A. Orr<sup>10</sup>, Rhamy Zeid<sup>1</sup>, Donald R. Polaski<sup>1</sup>, Maia Segura-Wang<sup>2</sup>, Sebastian M. Waszak<sup>2</sup>, David T. W. Jones<sup>3,11</sup>, Marcel Kool<sup>3,11</sup>, Volker Hovestadt<sup>6</sup>, Ivo Buchhalter<sup>12</sup>, Laura Sieber<sup>3</sup>, Pascal Johann<sup>3</sup>, Lukas Chavez<sup>3</sup>, Stefan Gröschel<sup>13</sup>, Marina Ryzhova<sup>14</sup>, Andrey Korshunov<sup>11,15</sup>, Wenbiao Chen<sup>5</sup>, Victor V. Chizhikov<sup>16</sup>, Kathleen J. Millen<sup>7,17</sup>, Vyacheslav Amstislavskiy<sup>8</sup>, Hans Lehrach<sup>8</sup>, Marie-Laure Yaspo<sup>8</sup>, Roland Eils<sup>12,18</sup>, Peter Lichter<sup>6,11</sup>, Jan O. Korbel<sup>2</sup>, Stefan M. Pfister<sup>3,11,19§</sup>, James E. Bradner<sup>1§</sup> & Paul A. Northcott<sup>3,4§</sup>

**Medulloblastoma is a highly malignant paediatric brain tumour, often inflicting devastating consequences on the developing child. Genomic studies have revealed four distinct molecular subgroups with divergent biology and clinical behaviour. An understanding of the regulatory circuitry governing the transcriptional landscapes of medulloblastoma subgroups, and how this relates to their respective developmental origins, is lacking. Here, using H3K27ac and BRD4 chromatin immunoprecipitation followed by sequencing (ChIP-seq) coupled with tissue-matched DNA methylation and transcriptome data, we describe the active *cis*-regulatory landscape across 28 primary medulloblastoma specimens. Analysis of differentially regulated enhancers and super-enhancers reinforced inter-subgroup heterogeneity and revealed novel, clinically relevant insights into medulloblastoma biology. Computational reconstruction of core regulatory circuitry identified a master set of transcription factors, validated by ChIP-seq, that is responsible for subgroup divergence, and implicates candidate cells of origin for Group 4. Our integrated analysis of enhancer elements in a large series of primary tumour samples reveals insights into *cis*-regulatory architecture, unrecognized dependencies, and cellular origins.**

Medulloblastoma is a highly malignant paediatric brain tumour classified into four biologically and clinically distinct molecular subgroups<sup>1</sup>. The present clinical approach to medulloblastoma involves maximal safe surgical resection, cytotoxic chemotherapy, and cranio-spinal radiation, which together are associated with profound morbidity in the developing child, underscoring the need for new subgroup-specific therapeutic insights.

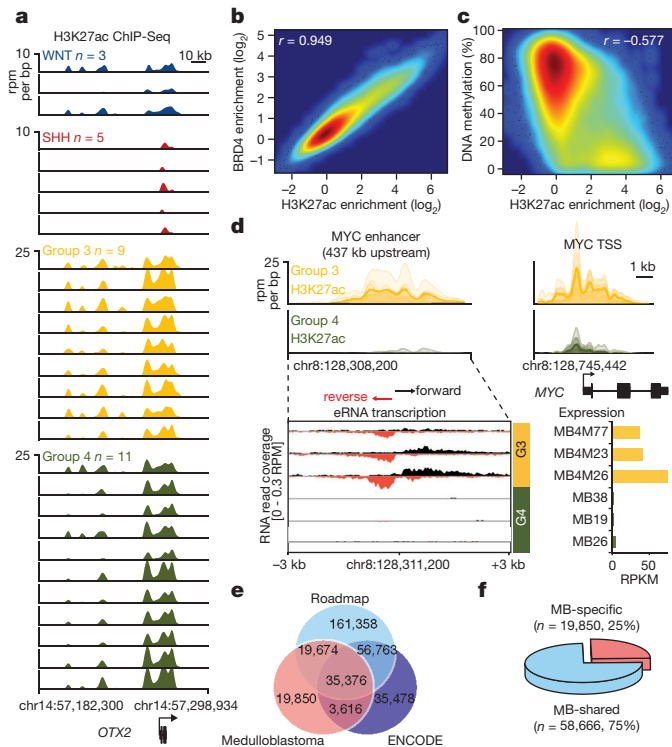
Transcriptional diversity amongst WNT, SHH, Group 3, and Group 4 subgroup medulloblastomas is partially explained by active and discriminatory signalling pathways, such as the Wingless/WNT and Sonic hedgehog/SHH developmental cascades inherent to WNT and SHH medulloblastomas, respectively. Somatic altered driver genes including *MYC* (Group 3), *KDM6A* (Group 4), *GFI1* and/or *GFI1B* (Group 3 and Group 4), and others contribute further to subgroup divergence<sup>2–4</sup>. Recurrent targeting of genes involved in chromatin modification has been the most consistent theme to emerge from recent next-generation sequencing (NGS) studies<sup>5</sup>, strongly suggesting deregulation of the epigenome as a critical step during medulloblastoma pathogenesis. However, this hypothesis has yet to be experimentally substantiated and knowledge pertaining to how the medulloblastoma epigenome influences subgroup-specific transcriptional programs remains in its infancy<sup>6</sup>.

Enhancers are *cis*-acting regulatory elements that recruit transcription factors (TFs) and chromatin-associated regulatory complexes, which together signal to RNA polymerase to regulate target gene expression<sup>7</sup>. Consortia such as ENCODE<sup>8,9</sup> and the Roadmap Epigenomics Consortium<sup>10</sup> have extensively mapped enhancers, advancing our understanding of enhancer/gene regulation across a comprehensive spectrum of cell lines and tissues. These resources empower our understanding of the complex cartography of the human regulatory landscape, provide testable hypotheses regarding disease-risk association, contribute evolutionary inferences, and establish robust analytical techniques. To deeply characterize the active *cis*-regulatory circuitry of a single disease entity, here medulloblastoma, we performed high-resolution chromatin immunoprecipitation with sequencing (ChIP-seq) for active enhancers (H3K27ac) in 28 primary tumour specimens and three established cell lines. Our approach to studying enhancers genome-wide in a large set of primary tissue samples led to a regulatory explanation for subgroup transcriptional diversity, previously unrecognized subgroup-specific dependencies, and firm insights into medulloblastoma cellular origins, in particular for the poorly characterized Group 3 and Group 4 subgroups.

<sup>1</sup>Medical Oncology, Dana Farber Cancer Institute (DFCI), Boston, Massachusetts 02215, USA. <sup>2</sup>Genome Biology Unit, European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany. <sup>3</sup>Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. <sup>4</sup>Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. <sup>5</sup>Department of Molecular Physiology & Biophysics, Vanderbilt University School of Medicine, Nashville, Tennessee 37212, USA. <sup>6</sup>Division of Molecular Genetics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. <sup>7</sup>Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, Washington 98105, USA. <sup>8</sup>Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. <sup>9</sup>Department of Bone Marrow Transplantation & Cellular Therapy, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. <sup>10</sup>Department of Pathology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. <sup>11</sup>German Cancer Consortium (DKTK), 69120 Heidelberg, Germany. <sup>12</sup>Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. <sup>13</sup>Department of Translational Oncology, NCT Heidelberg, 69120 Heidelberg, Germany. <sup>14</sup>Department of Neuropathology, NN Burdenko Neurosurgical Institute, 125047 Moscow, Russia. <sup>15</sup>Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), and Department of Neuropathology University Hospital, 69120 Heidelberg, Germany. <sup>16</sup>Department of Anatomy and Neurobiology, University of Tennessee Health Sciences Center, Memphis, Tennessee 38163, USA. <sup>17</sup>Department of Pediatrics, Genetics Division, University of Washington, Seattle, Washington 98195, USA. <sup>18</sup>Institute of Pharmacy and Molecular Biotechnology and BioQuant, University of Heidelberg, 69117 Heidelberg, Germany. <sup>19</sup>Department of Pediatrics, University of Heidelberg, 69117 Heidelberg, Germany. †Present address: Baylor College of Medicine, Department of Molecular and Human Genetics, Houston, Texas 77030, USA.

\*These authors contributed equally to this work.

§These authors jointly supervised this work.



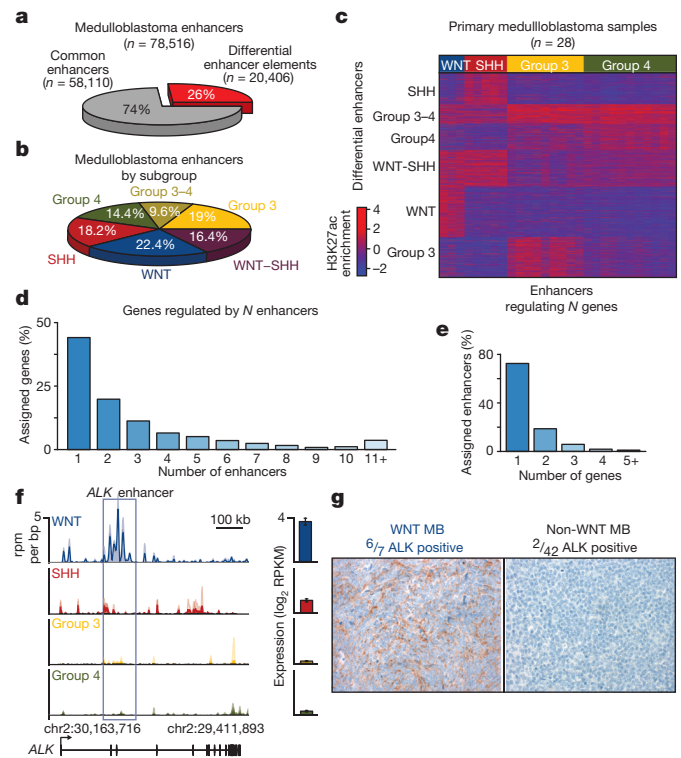
**Figure 1 | The enhancer landscape of primary medulloblastoma.**

**a**, Highly active enhancers at the *OTX2* locus across 28 primary medulloblastomas. **b**, H3K27ac versus BRD4 ChIP-seq signals at medulloblastoma enhancers ( $n = 78,516$ ). **c**, H3K27ac ChIP-seq signal versus DNA methylation (WGBS) at medulloblastoma enhancers ( $n = 78,516$ ). **d**, Group 3-specific eRNA expression (lower left) overlapping a Group 3-specific *MYC* enhancer (upper left) in a subset of medulloblastomas ( $n = 6$ ). *MYC* gene expression (reads per kilobase of transcript per million, RPKM) is also shown for the same cases (lower right). **e**, **f**, Overlap of medulloblastoma enhancers with ENCODE and Roadmap enhancers.

## The medulloblastoma enhancer landscape

Recent large-scale efforts annotating active regulatory elements genome-wide in human tissues (for example, through DNase I hypersensitivity, H3K27ac and BRD4 ChIP-seq), have catalogued enhancers in immortalized or malignant cell lines and normal human tissues, often under-representing discrete disease entities<sup>8,10</sup>. For medulloblastoma, only a single long-term culture cell line (D721; first reported in 1997) is included amongst 125 cell types initially studied by ENCODE<sup>9</sup>. Further, cancer cell lines often exhibit drastic genomic and transcriptional divergence from their corresponding primary tumour tissues as exemplified in non-Hodgkin's lymphoma where our prior epigenomic analyses identified greater likeness between primary tumour samples and normal lymphoid tissues than between tumours and cell lines<sup>11</sup>. Given the apparent limitations of using cell lines to faithfully study the tumour epigenome, and the recognized subgroup-dependent heterogeneity of medulloblastoma, we collected a series of 28 treatment-naïve, fresh-frozen medulloblastoma specimens and profiled the active enhancer landscape by H3K27ac ChIP-seq (Fig. 1a; Extended Data Fig. 1a–c).

This cohort is inclusive of all four medulloblastoma subgroups (Supplementary Table 1; WNT,  $n = 3$ ; SHH,  $n = 5$ ; Group 3,  $n = 9$ ; Group 4,  $n = 11$ ) and includes three additional Group 3 cell lines (MED8A, D425, and HD-MB03). Using MACS<sup>12</sup> to identify significantly enriched H3K27ac peaks, we inferred 78,516 enhancers, effectively saturating the medulloblastoma enhancer landscape (Extended Data Fig. 1d). These regions of promoter distal H3K27ac enrichment mainly (~80%) covered introns and intergenic regions (Extended Data Fig. 1e). Parallel ChIP-seq was performed for bromodomain



**Figure 2 | Differentially regulated enhancers in medulloblastoma subgroups.**

**a**, ANOVA classification of medulloblastoma enhancers. **b**, Distribution of differentially regulated enhancers among medulloblastoma enhancer classes. **c**, K-means clustering of differentially regulated medulloblastoma enhancers ( $n = 20,406$ ). **d**, Proportion of enhancer/gene assignments to *N* enhancers. **e**, Proportion of enhancer/gene assignments to *N* genes. **f**, WNT-specific enhancer activity (reads per million mapped reads per bp, rpm per bp) and expression ( $\log_2$  RPKM,  $n = 140$ ) of *ALK*. Error bars represent standard deviation (s.d.) of the mean. **g**, Immunohistochemical validation of ALK expression in WNT medulloblastoma patients ( $n = 49$ ).

containing 4 (BRD4), an enhancer-associated transcriptional coactivator<sup>11,13</sup>, in 27/31 cases. Enrichment of H3K27ac and BRD4 ChIP-seq signals strongly correlated at putative enhancer loci (Pearson correlation,  $r = 0.949$ ), further enforcing their active enhancer classification (Fig. 1b)<sup>11,13</sup>. Likewise, H3K27ac peaks were strongly anti-correlated with DNA methylation (Pearson correlation,  $r = -0.577$ ; Fig. 1c) and showed a high degree of overlap with the active/poised enhancer H3K4me1 but not the repressive H3K27me3 histone marks (Extended Data Fig. 1f). Finally, strand-specific RNA-seq data generated from the same cohort detected short, unspliced, bidirectional RNA transcripts overlapping H3K27ac peaks (Fig. 1d), in accordance with recently described enhancer RNAs (eRNAs)<sup>14</sup>. Active enhancers exhibited a modest statistical enrichment for overlap with focal amplifications and deletions identified in Group 3 and Group 4 (ref. 4;  $P = 0.028$  for amplifications,  $P = 0.016$  for deletions; Extended Data Fig. 1g). Comparison of predicted medulloblastoma enhancers with those reported using analogous methods employed by the ENCODE and Roadmap Epigenomics projects revealed 19,850 novel regulatory regions, indicative of potentially hindbrain- or medulloblastoma-specific enhancers in our data set (Fig. 1e, f). Primary medulloblastoma enhancer landscapes exhibited poor overlap and correlation with those generated from medulloblastoma cell lines (Extended Data Fig. 1h, i), further emphasizing the importance of studying the epigenome in primary tumours.

Analysis of variance (ANOVA) identified sets of enhancers differing according to known molecular subgroup, revealing 20,406 differentially active enhancers (26% of all inferred enhancers; Fig. 2a, b). The remaining 74% ( $n = 58,110$ ) displayed varied activity across



subgroups, suggesting either ubiquitous activity of for example, 'housekeeping' genes or a general role in medulloblastoma or cerebellar identity (Fig. 2a; Supplementary Table 2). *K*-means clustering of differentially regulated enhancers delineated six distinct medulloblastoma enhancer classes, including one for each subgroup as well as WNT-SHH and Group 3-Group 4 shared classes (Fig. 2b, c). Group 3 and Group 4 subgroups are known to exhibit some degree of transcriptional similarity<sup>15,16</sup>, consistent with the enhancer clustering results, whereas a common subset of shared enhancers between WNT and SHH subgroups was unexpected.

### Medulloblastoma enhancer/gene assignment

We next sought to assign enhancer elements to target genes, a process typically hindered by the majority of enhancer–promoter interactions occurring over extensive and highly variable genomic distances<sup>17</sup>. To overcome these challenges, we leveraged sample-matched RNA-seq gene expression data to identify putative enhancer/gene interactions that are (1) contained within the same topologically associated domain (TAD<sup>18</sup>) and (2) exhibit significant positive correlations between enhancer H3K27ac signal and gene expression (false discovery rate (FDR) < 0.05, Extended Data Fig. 2a–i). This approach assigned 8,775 enhancers (43% of all differential enhancers) to at least one protein-coding target gene (Supplementary Table 3). The majority (44%) of inferred target genes were assigned to a single enhancer, but in many cases, several enhancers were predicted to converge on the regulation of a single gene (Fig. 2d). Likewise, 73% of enhancers were assigned to only a single gene target (Fig. 2e). To validate the robustness of our methods, we used 4C-seq<sup>19</sup> to query Group 3-specific enhancer–promoter interactions for enhancers showing conserved activity in both primary Group 3 tumours and cell lines. This approach confirmed enhancer–promoter interactions for both *TGFBR1* and *SMAD9* in the Group 3 cell line HD-MB03, a low-passage line more faithful to primary Group 3 tumours than older models<sup>6,20</sup> (Extended Data Fig. 2j, k).

Medulloblastoma subgroup 'signature' genes have been extensively documented using various expression-profiling methods<sup>15,16</sup>. Enhancer/gene assignments derived from coupling H3K27ac ChIP-seq with RNA-seq produced a refined 'lens' for investigating subgroup-related diversity in medulloblastoma, implicating themes previously undisclosed through expression data alone. For example, enhancers

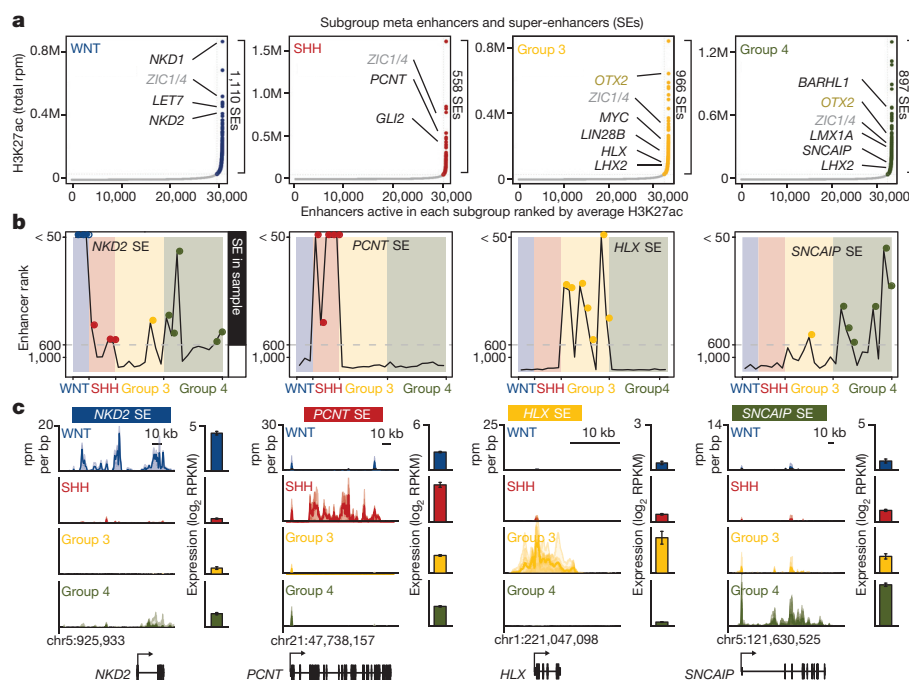
regulating *ALK*, a receptor tyrosine kinase frequently altered in a variety human cancers<sup>21</sup>, were found to be highly active in the WNT subgroup and explained the largely WNT-specific expression pattern detected by RNA-seq and confirmed by immunohistochemical staining of primary patient samples ( $n = 49$ ;  $P = 1.35 \times 10^{-5}$ , Fisher's exact test; Fig. 2f, g). Further investigations into the potential oncogenic role of *ALK* in WNT subgroup medulloblastoma are essential but rational given that *ALK* inhibitors are currently FDA-approved for the treatment of non-small-cell lung carcinoma (that is, crizotinib and ceritinib).

Rational target-based treatment options remain scarce for Group 3 and Group 4 subgroup patients, necessitating additional biological insights to direct future mechanistic and translational research. Functional pathway analysis (see Supplementary Methods) performed on differential enhancer/gene target assignments identified enrichment of neuronal transcriptional regulators in Group 4 and thematic pathways associated with TGF- $\beta$  signalling in Group 3 (Extended Data Fig. 3a–c). Notably, we uncovered a ~450 kilobase focal amplification at the *ACVR2A* locus in one Group 3 sample that encompassed both the gene and the upstream enhancer regions (Extended Data Fig. 3d). In this sample, enhancers regulating TGF- $\beta$  pathway components exhibited increased H3K27ac versus other Group 3 tumours (Extended Data Fig. 3e). These data, combined with our prior observations that TGF- $\beta$  receptor genes are recurrently amplified in Group 3 (ref. 4), further suggest TGF- $\beta$  signalling as a putative oncogenic driver in this subgroup.

### Medulloblastoma subgroup super-enhancers

In multiple tumour types, super-enhancers (SEs), broad spatially co-localized enhancer domains<sup>22,23</sup>, have recently been shown to drive oncogenes, genes required for maintenance of tumour cell identity, and genes associated with cell-type-specific functions. To determine whether SEs might play a role in characterizing subgroup-specific identity, we undertook a systematic mapping of SEs across all 28 medulloblastoma samples (see Supplementary Methods; Extended Data Fig. 4a). Massive (>50 kb) SE domains were identified at the cerebellar-specific TFs, *ZIC1* and *ZIC4*<sup>24</sup> (Extended Data Fig. 4b, c), and at ~70% of a queried set of established medulloblastoma driver genes and chromatin modifiers implicated in cancer, including *GLI2*, *MYC*, *OTX2* and others<sup>4</sup> (Extended Data Fig. 4d).

To identify subgroup-specific SEs, we took the union of all enhancer regions in a given subgroup and ranked them by average H3K27ac



**Figure 3 | Medulloblastoma super-enhancers characterize subgroup-specific identity.**

**a**, Ranked enhancer plots defined across composite H3K27ac landscapes of WNT, SHH, Group 3, and Group 4 medulloblastomas. Select genes associated with SEs in each subgroup are highlighted and shaded according to enhancer class specificity. **b**, Enhancer rankings for candidate subgroup-specific SEs across all samples according to subgroup. **c**, Meta tracks of H3K27ac ChIP-seq signal (rpm/bp) across medulloblastoma subgroups for the loci shown in **b**. Candidate gene expression (mean RPKM) is shown to the right of each H3K27ac track ( $n = 140$ ). Error bars represent standard deviation (s.d.) of the mean.

enrichment across all samples in that subgroup<sup>23</sup>, resulting in ~3,000 distinct SE-containing loci (~600–1,100 SEs per subgroup; Fig. 3a; Extended Data Fig. 4e; Supplementary Table 4). Compared to typical enhancers, SEs showed higher occupancy of BRD4 and greater enhancer signal dynamic range between subgroups (Extended Data Fig. 4f–h). Targets of differential enhancers contained within SEs (that is, SE target genes) included a large fraction of established medulloblastoma signature genes (32%; Supplementary Table 3), as well as novel candidates (Fig. 3a–c). Medulloblastoma SEs were inferred to regulate known Cancer Gene Census genes, including the aforementioned *ALK* in WNT, *SMO* and *NTRK3* in SHH, *LMO1*, *LMO2*, and *MYC* in Group 3, and *ETV4* and *PAX5* in Group 4, among others (Supplementary Table 3). Furthermore, several actionable, SE-regulated genes were revealed in our analysis (Supplementary Table 5).

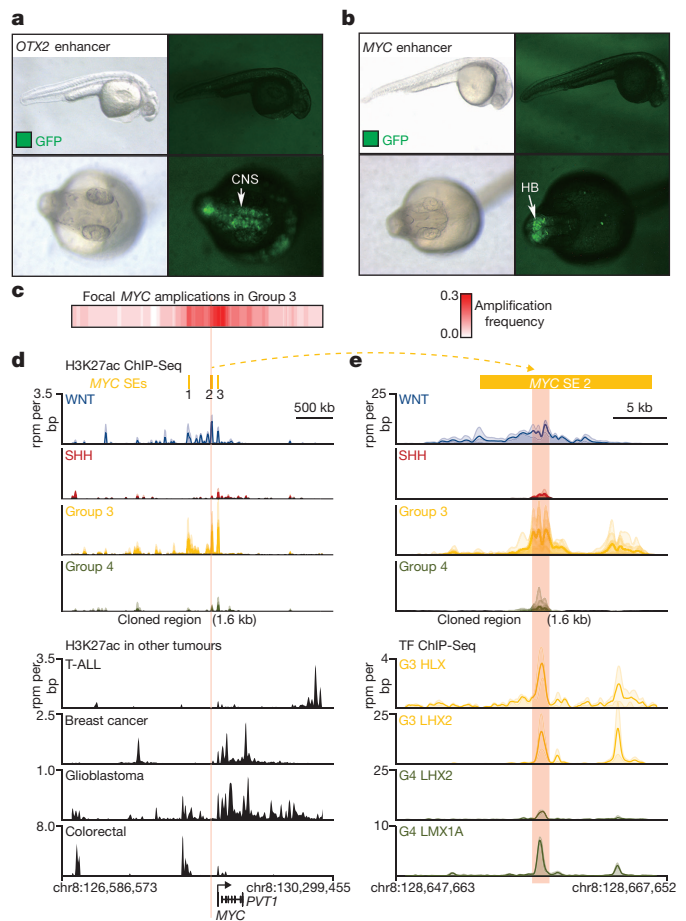
Unbiased hierarchical clustering of SEs across samples was sufficient to recapitulate transcriptional subgroupings using no prior knowledge of subgroup status, suggesting that SEs might play a pivotal role in characterizing subgroup identity (Extended Data Fig. 4a). SEs from established Group 3 medulloblastoma cell lines clustered with one another, but failed to show similarity to primary Group 3 samples or samples from any other subgroup.

To experimentally validate the activity of medulloblastoma subgroup-specific SEs, we synthesized twenty-two unique SE loci (size range, 1.1–2.1 kb) and evaluated them using Tol2 transposon-mediated zebrafish transgenesis (see Supplementary Methods)<sup>25</sup>. These *in vivo* reporter assays resulted in a validation rate of 45% (10/22), with all reproducibly active enhancer constructs showing specific activity in the zebrafish central nervous system (Fig. 4a, b; Extended Data Fig. 5a–l). We used TF ChIP-seq data for HLX, LHX2, and LMX1A—all highly expressed and SE-regulated in Group 3 and/or Group 4 (Fig. 3a and data not shown)—to enable precise definition of enhancer coordinates (based on TF occupancy) for testing in zebrafish (Fig. 4e), potentially explaining the remarkably high *in vivo* validation rate we observed. These experiments confirmed zebrafish hindbrain-specific activity for an SE (active in WNT and Group 3) mapping approximately 90 kb upstream of *MYC* inferred to regulate *MYC* expression (Fig. 4b–e). This SE was not found in other common human cancers (Fig. 4d), and in only 4/77 different primary tissues included in Roadmap, suggesting that this validated *MYC* SE is highly specific to the developing hindbrain and/or medulloblastoma (Extended Data Fig. 5m). Importantly, identified *MYC* SEs clearly demarcate a focal amplification hotspot in published Group 3 medulloblastoma copy-number data<sup>4</sup> (Fig. 4c, d), strongly implicating these SEs in the oncogenic regulation of *MYC*. Collectively, these *in vivo* validation data further substantiate our highly integrative approach for the identification of enhancers and SEs, and inference of their target genes.

### SE-regulated TFs reveal cellular origins

Among subgroup-specific SE target genes, we observed an enrichment of TFs involved in neuronal development ( $P \approx 0.0001$ , Fisher's exact test; Extended Data Fig. 6a). Overall, subgroup-specific TFs displayed similar patterns of expression, enhancer motif enrichment, and overlap of target genes (Extended Data Figs 6b and 7). TFs were also enriched in subgroup-specific SE targets as compared to subgroup-specific non-SE targets ( $P \approx 0.002$ , Fisher's exact test), consistent with prior observations that SEs regulate key TFs required for tumour cell identity and maintenance<sup>11,13,22</sup>. Given evidence in embryonic stem cells that pluripotency master regulator TFs (OCT4, SOX2, and NANOG) are driven by SEs and themselves bind to and establish SEs<sup>23</sup>, we hypothesized that a reverse analysis of SEs in medulloblastoma might enable a *de novo* reconstruction of tumour identity-defining TFs and their associated regulatory circuitry, thereby providing novel insights into medulloblastoma origins.

Pursuant to this idea, we proposed a definition of core regulatory circuitry TFs in which the TFs are SE-regulated and the TFs themselves bind to the SEs of one another (Fig. 5a, see Supplementary Methods). For each SE-regulated TF, these criteria are quantified by measuring the



**Figure 4 | *In vivo* validation of medulloblastoma super-enhancers.** **a, b**, Zebrafish reporter assays for *OTX2* (**a**) and *MYC* (**b**) enhancers observed in medulloblastoma. Arrows indicate the locations of GFP signal. CNS, central nervous system; HB, hindbrain. **c**, Heat map summarizing *MYC* copy-number data derived from a published series of Group 3 medulloblastomas ( $n = 168$ ). **d, e**, H3K27ac ChIP-seq (upper panels) data showing a shared WNT/Group 3 *MYC* enhancer not found in other human cancers (**d**, lower panel) and occupied by SE-regulated TFs HLX, LHX2, and LMX1A as determined by TF ChIP-seq (**e**, lower panel).

inward binding of other SE associated TFs (in degree) and the outward binding of the TF to other SEs (out degree) (Fig. 5a, b). Regulatory circuitry reconstruction across all SE-associated TFs in medulloblastoma identified cliques of TFs with similar patterns of in/out degree, strong interconnectivity via motif binding, and higher likelihoods of pairwise protein–protein interaction and motif co-occurrence at enhancers (see Supplementary Methods, Extended Data Fig. 8). This reconstruction creates for the first time a core regulatory circuitry blueprint for each subgroup, and implicates specific sets of TFs in establishing medulloblastoma subgroup identity (Extended Data Fig. 9). Importantly, ChIP-seq for the homeodomain TFs HLX (Group 3 network), LMX1A (Group 4 network), and LHX2 (shared Group 3/Group 4 network) performed on select Group 3 and Group 4 primary samples ( $n = 4$ ) largely validated the computationally derived regulatory networks constructed for these subgroups (Fig. 5c, d; Extended Data Figs 8 and 9).

Distinct cellular origins for WNT and SHH medulloblastomas have been experimentally established using a variety of genetically engineered mouse models<sup>26–28</sup>. The origins of Group 3 and Group 4 medulloblastoma, however, are unknown and yet essential to define, as these tumours account for ~60% of all diagnoses, lack targeted therapies, and are frequently associated with poor clinical outcomes<sup>1</sup>.

Cell identity is most essentially defined by the activity of master regulator TFs. As such, we hypothesized that the regulatory SE





**Figure 5 | Super-enhancers characterize medulloblastoma regulatory circuitry.** **a**, Methodology for inferring medulloblastoma core regulatory circuitry. **b**, Heat map of all SE-associated TFs in medulloblastoma (rows) clustered by similarity of regulatory degree. Selected TFs with similar subgroup-specific patterns of regulatory degree are annotated. **c**, Subgroup-specific regulatory circuitry in Group 3 and Group 4 medulloblastoma. **d**, TF and H3K27ac ChIP-seq meta tracks for the SE-regulated TFs *LMX1A*, *LHX2*, *HLX*, and *EOMES*.

regions governing endogenous expression of candidate master TFs and embedded in the core regulatory circuitry of medulloblastoma subgroups might inform cellular origins of the disease via their cell-type-specific activity. During early cerebellar development, *LMX1A*, *EOMES*, and *LHX2*—master regulator Group 4 TFs deduced from our core regulatory circuitry analysis (Fig. 5c)—exhibit overlapping spatiotemporal restricted expression in the nuclear transitory zone (NTZ; Fig. 6a), an assembly point for immature deep cerebellar nuclei (DCN). DCN residing in the NTZ at this time point are predominantly glutamatergic projection neurons that originate from earlier progenitors of the upper rhombic lip (uRL), a transient germinal zone producing progenitors with distinct cellular fates, including DCN and cerebellar granule neurons<sup>29</sup>. Immunofluorescence microscopy confirmed compartmentalized expression of *LMX1A*, *EOMES*, and

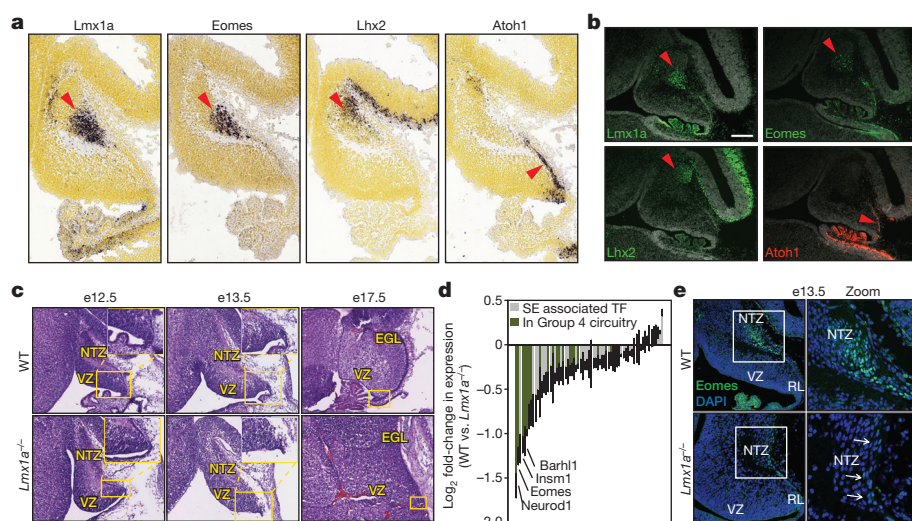
*LHX2* that was notably distinct from *ATOH1* expression, the latter marking the early external granule layer (EGL) at this developmental stage (Fig. 6b).

Both *LMX1A* enhancer activity and expression are highly discriminatory for Group 4, nominating this TF as a master regulator of the Group 4 transcriptional program (Figs 3a and 5c; Extended Data Figs 8 and 9). Indeed, *LMX1A* ChIP-seq performed on Group 4 primary samples verified >90% of predicted target genes inferred through motif-driven computational analyses (Extended Data Figs 8 and 9). *LMX1A* is a LIM-homeodomain TF previously shown to function as a critical regulator of cell-fate decisions in the uRL and essential for normal cerebellar development<sup>30</sup>. Spontaneous *Lmx1a* loss-of-function null mutations are causative in *dreher* mice, resulting in profound cerebellar phenotypes typified by premature regression of the RL, reduced choroid plexus, and cerebellar hypoplasia predominantly affecting the posterior vermis (Fig. 6c)<sup>31</sup>. To further investigate the molecular targets associated with *dreher* cerebellar phenotypes, we microdissected uRL from wild-type and *dreher* (*dr<sup>l</sup>/dr<sup>l</sup>*) mice at embryonic day 13.5 and delineated transcriptional differences through expression profiling. Strikingly, SE-regulated TFs contained in Group 3/Group 4 regulatory circuitry (Extended Data Fig. 9) were among the most differentially expressed genes in *dreher* uRL compared to controls (Fig. 6d, e). Collectively, these phenotypic and molecular data further support *Lmx1a* as a master regulator TF in both the cerebellar uRL and in Group 4 medulloblastoma, implicating the uRL compartment and its derivative precursors as putative cells-of-origin for Group 4.

## Discussion

We describe the active medulloblastoma enhancer landscape across a series of 28 fresh-frozen, treatment-naïve tissue samples and three cultured cell lines, to our knowledge representing the largest such data set for any single cancer entity. Our data reveal drastic divergence between primary tumour and tumour cell line material and uncover considerable *cis*-regulatory element heterogeneity between subgroups of the disease that would be unsubstantiated in series limited to just a few cases.

Clinically relevant medulloblastoma subgroups are principally defined based on their underlying transcriptional profiles. Differentially regulated medulloblastoma enhancers and SEs are here shown to recapitulate these subgroups, and importantly extend our understanding of this disease to inferences regarding cell specification and actionable tumour dependencies. Biological themes and signalling networks extracted from transcriptional data have served as the primary source of annotation for medulloblastoma subgroups, with WNT and SHH subgroups characterized by activation of their respective signalling pathways, and Group 3 and Group 4 recognized for their GABAergic and glutamatergic expression phenotypes, respectively. Although these



**Figure 6 | Master transcription factors implicate Group 4 cellular origins.**

**a**, Expression (*in situ* hybridization) of *Lmx1a*, *Eomes*, *Lhx2*, and *Atoh1* in the embryonic cerebellum (e13.5). **b**, Immunofluorescence microscopy for the TFs shown in **a** performed on sagittal sections of the e13.5 murine cerebellum. **c**, Haematoxylin and eosin-stained cerebellar sections (sagittal) of wild-type and *dr<sup>l</sup>/dr<sup>l</sup>* (*Lmx1a*<sup>-/-</sup>) embryos. The RL is demarcated by a yellow box in each panel. **d**, Differentially expressed TFs in the e13.5 RL of *Lmx1a*<sup>-/-</sup> embryos versus wild-type controls. Error bars represent standard error of the mean ( $n = 3$ ). **e**, Immunofluorescence microscopy confirming *Eomes* downregulation in *Lmx1a*<sup>-/-</sup> embryos (e13.5).



data provide a functional and phenotypic annotation of medulloblastoma, they fail to articulate the developmental identity of individual subgroups. Using a reverse analysis of the medulloblastoma chromatin landscape starting at the level of differentially regulated enhancers and SEs, we have reconstructed and experimentally validated the core regulatory circuitry inherent to medulloblastoma subgroups, inferring master transcriptional regulators responsible for subgroup-specific divergence. The majority of these master regulator TFs were not previously implicated in medulloblastoma development, nor were they visible amongst transcriptionally-derived gene sets dominated by overwhelming phenotypic signatures. Through tracing the spatiotemporal activity of a subset of Group 4 master TFs, these studies identified DCN of the cerebellar NTZ, or plausibly their earlier precursors originating from the uRL, as putative cells-of-origin for this large subgroup of patients. Together these approaches establish a framework for the inference of tumour cell-of-origin through enhancer core regulatory circuitry mapping.

Identifying the cellular origins of cancer has broad implications for the understanding and treatment of malignancy<sup>32</sup>. Although tumour cells deviate from their developmental origins during transformation, numerous cancers, especially those of the immune compartment, still maintain developmental TF activity and as such are treatable through targeting of the lineage (for example, anti-B cell therapies for leukaemia)<sup>33,34</sup>. As medulloblastoma is believed to originate from cell populations that normally exist ephemerally during development, targeting the aberrant persistence of tumour cells from these lineages may represent a novel therapeutic strategy with minimal effect on the normal tissue compartment. Moreover, elucidation of master TFs of medulloblastoma implicates upstream signalling pathways, transcriptional co-activators, and downstream effectors as potential subgroup-specific targets for rational therapeutic intervention. These insights demonstrate the critical importance of epigenetic analyses of primary tumours as opposed to cell line model systems and highlight the broad utility of core regulatory circuitry mapping, especially in poorly characterized and clinically heterogeneous malignancies.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 24 March; accepted 14 December 2015.**

**Published online 27 January 2016.**

- Northcott, P. A., Korshunov, A., Pfister, S. M. & Taylor, M. D. The clinical implications of medulloblastoma subgroups. *Nat. Rev. Neurol.* **8**, 340–351 (2012).
- Jones, D. T. et al. Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105 (2012).
- Northcott, P. A. et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).
- Northcott, P. A. et al. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488**, 49–56 (2012).
- Northcott, P. A. et al. Medulloblastomics: the end of the beginning. *Nature Rev. Cancer* **12**, 818–834 (2012).
- Hovestadt, V. et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).
- Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nature Rev. Genet.* **15**, 272–286 (2014).
- Consortium, E. P.; ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
- Kundaje, A. et al.; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- Chapuy, B. et al. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790 (2013).
- Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
- Lovén, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
- Kim, T. K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).

- Cho, Y. J. et al. Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J. Clin. Oncol.* **29**, 1424–1430 (2011).
- Northcott, P. A. et al. Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* **29**, 1408–1414 (2011).
- Jin, F. et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
- Pope, B. D. et al. Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–405 (2014).
- Gröschel, S. et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369–381 (2014).
- Milde, T. et al. HD-MB03 is a novel Group 3 medulloblastoma model demonstrating sensitivity to histone deacetylase inhibitor treatment. *J. Neurooncol.* **110**, 335–348 (2012).
- Hallberg, B. & Palmer, R. H. Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. *Nature Rev. Cancer* **13**, 685–700 (2013).
- Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
- Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
- Aruga, J. et al. Mouse Zic1 is involved in cerebellar development. *J. Neurosci.* **18**, 284–293 (1998).
- Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
- Gibson, P. et al. Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468**, 1095–1099 (2010).
- Schüller, U. et al. Acquisition of granule neuron precursor identity is a critical determinant of progenitor cell competence to form Shh-induced medulloblastoma. *Cancer Cell* **14**, 123–134 (2008).
- Yang, Z. J. et al. Medulloblastoma can be initiated by deletion of Patched in lineage-restricted progenitors or stem cells. *Cancer Cell* **14**, 135–145 (2008).
- Fink, A. J. et al. Development of the deep cerebellar nuclei: transcription factors and cell migration from the rhombic lip. *J. Neurosci.* **26**, 3066–3076 (2006).
- Chizhikov, V. V. et al. Lmx1a regulates fates and location of cells originating from the cerebellar rhombic lip and telencephalic cortical hem. *Proc. Natl Acad. Sci. USA* **107**, 10725–10730 (2010).
- Millonig, J. H., Millen, K. J. & Hatten, M. E. The mouse *Dreher* gene *Lmx1a* controls formation of the roof plate in the vertebrate CNS. *Nature* **403**, 764–769 (2000).
- Gilbertson, R. J. Mapping cancer origins. *Cell* **145**, 25–29 (2011).
- Byrd, J. C. et al. Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. *N. Engl. J. Med.* **369**, 32–42 (2013).
- Hale, G. et al. Remission induction in non-Hodgkin lymphoma with reshaped human monoclonal antibody CAMPATH-1H. *Lancet* **332**, 1394–1399 (1988).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** S.E. is a recipient of Human Frontiers Science Program long-term postdoctoral fellowship (LT000432/2014). S.M.W. received funding through a SNSF Early Postdoc Mobility Fellowship (P2ELP3\_155365) and an EMBO Long-Term Fellowship (ALTF 755-2014). C.Y.L. is supported by a US Department of Defense CDMRP CA120184 postdoctoral fellowship. P.A.N. is a V Foundation V Scholar in Childhood Cancer Research. We thank Creative Science Studios (<http://www.creativesciencesstudios.com/>) for assistance with artwork. We acknowledge J. Zhang and X. Zhou for their assistance in making this data accessible via the St. Jude PeCan Data Portal (<https://pecan.stjude.org/dataset/northcott>).

**Author Contributions** P.A.N., J.E.B., and S.M.P. conceived and co-led the study. C.Y.L. and S.E. performed all bioinformatics related to the analysis of medulloblastoma enhancers and super-enhancers. Y.T., L.Y., D.K., B.C.W., B.J., and W.C. validated subgroup-specific enhancers *in vivo*. C.Y.L. and A.J.F. constructed medulloblastoma regulatory circuitry networks. M.Z., S.M.W., R.Z., D.R.P., M.S.-W., D.T.W.J., M.K., V.H., I.B., and L.C. provided informatics and general scientific support. P.H., V.V.C., and K.J.M. performed the developmental studies with *dreher* and WT mouse embryonic cerebella. T.R., H.-J.W., V.A., H.L., and M.-L.Y. conducted RNA-seq data generation and enhancer RNA analysis. B.A.O. performed ALK staining on medulloblastoma TMA. L.S., P.J., and S.G. performed 4C-seq experiments. M.R. and A.K. provided medulloblastoma tissue samples. R.E., P.L., J.O.K., S.M.P., J.E.B., and P.A.N. provided institutional support and project supervision. C.Y.L., S.E., S.M.P., J.E.B., and P.A.N. prepared the figures and wrote the manuscript.

**Author Information** Short-read sequencing data have been deposited at the European Genome-Phenome Archive (EGA, <http://www.ebi.ac.uk/ega/>) hosted by the EBI, under accession number EGAS00001000215. The data generated in this study are also available at the St. Jude PeCan Data Portal (<https://pecan.stjude.org/dataset/northcott>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.A.N. ([paul.northcott@stjude.org](mailto:paul.northcott@stjude.org)), J.E.B. ([james\\_bradner@dfci.harvard.edu](mailto:james_bradner@dfci.harvard.edu)), and S.M.P. ([s.pfister@dkfz-heidelberg.de](mailto:s.pfister@dkfz-heidelberg.de)).

## METHODS

No statistical methods were used to predetermine sample size.

**Identifying super-enhancer constituents for reporter assays.** We sought to identify candidate Group 3 and Group 4 super-enhancer constituents for validation by reporter assays. We identified candidate Group 3 and Group 4 super-enhancer constituents by first locating nucleosome free “valleys” in the H3K27ac data using an algorithm adapted from ref. 35. Valleys that showed strong evidence of TF ChIP-seq binding for respective Group 3 (HLX and LHX2) and Group 4 (LHX2 and LMX1A) TFs were selected and manually curated for validation in reporter assays. Based on restrictions for DNA synthesis and cloning, candidate reporter regions of roughly  $\pm 1$  kb flanking the valley centre were used (Fig. 4 and Extended Data Fig. 5).

**Zebrafish *in vivo* enhancer assays.** All experiments involving zebrafish (*Danio rerio*, AB strain) were approved by the Vanderbilt Institutional Animal Care and Use Committee. For *in vivo* zebrafish reporter assays, a minimum, ~150–200 embryos (male and female) were injected per reporter construct and assays were repeated 2–3 times per construct to confirm reproducibility. No randomization of enhancer assays was performed. The scientist who performed the injections had no prior knowledge related to the enhancer constructs and was therefore blinded to the experiment. Microinjection was done as described previously<sup>36</sup>. In brief, a mixture of individual enhancer-containing vector DNA (25  $\mu$ g ml<sup>-1</sup>) and transposase RNA (25  $\mu$ g ml<sup>-1</sup>) was injected into zebrafish zygotes (1 nl per zygote). The injected embryos were cultured in 0.3 $\times$  Danieau's solution at 28.5°C. After 24 hours, the embryos were examined for eGFP expression under a fluorescent dissecting microscope (Zeiss Discovery V12) to determine the stereotypic expression pattern conferred by the enhancer. The total number of embryos injected with the construct and the number of embryos with the stereotypical eGFP pattern were determined to calculate the frequency of the pattern. Embryos were decolorized and imaged using a Zeiss AxioCam HRC digital camera.

**Immunofluorescence microscopy.** Spatial protein expression of medulloblastoma<sub>G4</sub>-specific transcription factors in e13.5 cerebella was determined by IHC. PFA-fixed frozen tissues were sectioned (12  $\mu$ m thickness) and processed without antigen retrieval steps. The antibodies used here are Tbr2 (1:100, Abcam, ab23345), Lmx1a (1:100, Novus Biologicals, NBP1-81303), Atoh1 (1:500, Abcam, ab105497) and appropriate secondary antibodies conjugated with Alexa fluorophores (1:400, Invitrogen). The images were captured by an epifluorescence microscopy.

**Analysis of Allen Brain Atlas data portal.** Endogenous expression of candidate TFs was determined by querying the Allen Brain Atlas Data Portal (<http://developingmouse.brain-map.org>) at various developmental time points.

**Medulloblastoma tissue microarrays (TMAs).** The molecular subgroup of 49 medulloblastoma samples on tissue microarrays were determined as previously described<sup>37</sup>. Immunohistochemistry was performed using clone ALK01 (#790-2918, Ventana) with appropriate secondary reagents. Individual tumours were scored positive in the presence of cytoplasmic immunoreactivity for ALK1, whereas the tumour was considered negative in the absence of immunoreactivity.

**Phenotypic analysis of Dreher (*Lmx1a*<sup>-/-</sup>) embryos.** All mouse (*Mus musculus*, B6C3HFe background) experiments were done in accordance with the guidelines laid down by the Institutional Animal Care and Use Committee (IACUC), of Seattle Children's Research Institute. No randomization or experimental blinding related to mouse experiments was performed. *Lmx1a*<sup>+/-</sup> mice were crossed and the day of plug was taken as e0.5. WT and *Lmx1a*<sup>-/-</sup> embryos (male and female) were dissected out between e12.5 and e17.5 and subsequently fixed in 4% paraformaldehyde (PFA) for 2–6 hours. The fixed embryos were washed in PBS and incubated in 30% sucrose overnight. The following day, embryos were frozen in optimum cutting temperature (OCT) compound. Mid-sagittal cryosections of the cerebellum at 11  $\mu$ m were taken. Haematoxylin and eosin staining and immunohistochemistry were performed as described previously<sup>38</sup>. Briefly, cryosections were incubated at room temperature for 1 hour after which they were subjected to heat-mediated antigen retrieval. All sections were blocked using 5% serum containing 0.35% Triton X, and then incubated with the primary antibody (Eomes (Tbr2); #14-4875, ebioscience, mouse, 1:200), overnight. The following day fluorescent-dye-labelled secondary antibodies (Alexa Fluor 488, 1:1000, Molecular Probes) were used. Sections were counter stained using DAPI (4',6'-diamidino-2-phenylindole) (Vector Laboratories). All images were captured at room temperature. Haematoxylin and eosin-stained sections were imaged using Hamamatsu Nanoscope whole slide scanner. All confocal images were captured using Zeiss LSM Meta and Zen 2009 software.

**Collection of patient material and cell lines.** An Institutional Review Board ethical vote (Ethics Committee of the Medical Faculty of Heidelberg) was obtained according to ICGC guidelines (<http://www.icgc.org>), along with informed consent for all participants. No patient underwent chemotherapy or radiotherapy before surgical removal of the primary tumour. Tumour tissues were subjected

to neuropathological review for confirmation of histology and for tumour cell content >80%. The ChIP-seq cohort was established based on tissue availability and availability of orthogonal data types (for example, WGS, RNA-seq) and patient metadata (for example, molecular subgroup). Subgroup assignments were made using the Illumina 450K DNA methylation array as described<sup>39</sup>. Medulloblastoma cell lines were cultured at 37°C with 5% CO<sub>2</sub>. D425\_Med (D425; a gift from D. D. Bigner) and MED8A cells (from the authors' own stocks; T. Pietsch) were cultured in DMEM with 10% FCS (Life Technologies). HD-MB03 cells<sup>20</sup> were grown in RPMI-1640 with 10% FCS (Life Technologies). All cells were regularly authenticated and tested for mycoplasma (Multiplexion, Heidelberg, Germany).

**ChIP-seq.** H3K27ac, BRD4, H3K27me3, H3K4me1, LMX1A, LHX2, and HLX ChIP was performed at ActiveMotif (Carlsbad, CA) using antibodies against H3K27ac (AM#39133, Active Motif), BRD4 (#A301-985A, Bethyl Laboratories), H3K27me3 (#07-449, Millipore), H3K4me1 (AM#39298, ActiveMotif), LMX1A (#AB10533, Millipore), LHX2 (#sc-19344, Santa Cruz), and HLX (#HPA005968, Sigma). Fresh-frozen medulloblastoma tissues (or cell lines) were submersed in PBS + 1% formaldehyde, cut into small pieces and incubated at room temperature for 15 min. Fixation was stopped by the addition of 0.125 M glycine (final concentration). The tissue pieces were then treated with a TissueTearer and finally spun down and washed 2 $\times$  in PBS. Chromatin was isolated by the addition of lysis buffer, followed by disruption with a Dounce homogenizer. Lysates were sonicated and the DNA sheared to an average length of ~300–500 bp. Genomic DNA (input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking, followed by ethanol precipitation. Pellets were resuspended and the resulting DNA was quantified on a NanoDrop spectrophotometer. Extrapolation to the original chromatin volume allowed quantitation of the total chromatin yield. An aliquot of chromatin (30  $\mu$ g) was precleared with protein A (G – for goat pc or monoclonal antibodies) agarose beads (Invitrogen). Genomic DNA regions of interest were isolated using 4  $\mu$ g of antibody. ChIP complexes were washed, eluted from the beads with SDS buffer, and subjected to RNase and proteinase K treatment. Crosslinks were reversed by incubation overnight at 65°C, and ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation. Quantitative PCR (qPCR) reactions were carried out in triplicate on specific genomic regions using SYBR Green Supermix (Bio-Rad). The resulting signals were normalized for primer efficiency by carrying out qPCR for each primer pair using Input DNA.

Illumina sequencing libraries were prepared from the ChIP and Input DNAs by the standard consecutive enzymatic steps of end-polishing, dA-addition, and adaptor ligation. After a final PCR amplification step, the resulting DNA libraries were quantified and sequenced on the Illumina HiSeq 2000 platform using 2 $\times$  101 cycles according to the manufacturer's instructions. Alignment, and downstream processing of ChIP-seq data was performed as described<sup>6</sup>.

**RNA-seq and transcriptome read alignment.** RNA was extracted from fresh frozen tissue samples using the AllPrep DNA/RNA/Protein Mini kit (Qiagen) including DNase I treatment on column. All samples were subjected to quality control on a Bioanalyzer instrument. RNA sequencing libraries were prepared from 10  $\mu$ g of total RNA. Strand-specific RNA sequencing was performed following a protocol described previously<sup>40,41</sup>. Sequencing was carried out with 2 $\times$  51 cycles on a HiSeq 2000 instrument (Illumina). All reads were aligned to the human reference genome (1000 genomes version of human reference genome hg19/GRCh37) using BWA (v 0.5.9-r16). Aligned reads were converted to the SAM/BAM format using SAMtools. Gene annotation was based on Ensembl v70 (*Homo sapiens*).

**4C-seq.** 4C samples were prepared from Group 3 medulloblastoma cell line HD-MB03 using the method as described<sup>19,42</sup>. DpnII was used as the primary restriction enzyme and Csp6I as the secondary restriction enzyme in template generation. Sample libraries for SMAD9 and TGFBR1 were amplified using the primers, SMAD9\_F: TTATCCAGGCAAGGAAGATC, SMAD9\_R: ATTACCTCATCTGCAAAACC, TGFBR1\_F: CATTCTTCTCCCCATGATC, and TGFBR1\_R: ACACAATCTTGGGTGTTTTT, respectively. Amplified libraries were multiplexed, spiked with 40% PhiX viral genome and sequenced on HiSeq 2000. Reads were mapped to human genome (hg19) using Bowtie (v 1.0.0)<sup>43</sup>.

**Identification of enhancer RNA candidates.** Forward and reverse RNA transcription based on directional RNA sequencing data was quantified in 3 kb windows upstream and downstream of enhancer peaks that were based on H3K27ac ChIP-seq data, resulting in four RNA expression values for each enhancer region: (L<sub>fwd</sub>) forward transcription left of enhancer peak, (R<sub>fwd</sub>) forward transcription right of enhancer peak, (L<sub>rev</sub>) reverse transcription left of enhancer peak, and (R<sub>rev</sub>) reverse transcription right of enhancer peak. We calculated the “directionality index” D, a measure of the directionality of transcription inside an enhancer region, with D ranging from 0 to 1, by  $D = |R_{fwd} - L_{rev}| / (R_{fwd} + L_{rev})$  as described before<sup>14</sup>, with low D values representing bidirectional eRNA transcription.



For correlation of eRNA transcription values with corresponding gene expression values, we calculated eRNA transcription values in 3 kb windows upstream and downstream of enhancer peaks by  $eRNA\_transcription = (R\_fwd + L\_rev)/2$ .

**Genomic coordinates and gene annotation.** All coordinates in this study were based on human reference genome assembly hg19, GRCh37 (<http://www.ncbi.nlm.nih.gov/assembly/2758/>). Gene annotations were based on gencode annotation release 19 (<http://www.gencodegenes.org/releases/19.html>).

**Calculating read density.** We calculated the normalized read density of a ChIP-seq data set in any genomic region using the Bamliquidator (version 1.0) read density calculator (<https://github.com/BradnerLab/pipeline/wiki/bamliquidator>). Briefly, ChIP-seq reads aligning to the region were extended by 200 bp and the density of reads per base pair (bp) was calculated. The density of reads in each region was normalized to the total number of million mapped reads producing read density in units of reads per million mapped reads per bp (rpm per bp).

**Plotting meta representations of ChIP-seq signal.** To compactly display medulloblastoma H3K27ac ChIP-seq signal at individual genomic loci and across subgroups, we developed a simple meta representation (Fig. 1d and others). For all samples within a group, ChIP-seq signal is smoothed using a simple spline function and plotted as a translucent shape in units of rpm per bp. Darker regions indicate regions with signal in more samples. An opaque line is plotted and gives the average signal across all samples in a group.

**Peak finding and classification.** H3K27ac peak finding was performed using MACS<sup>12</sup> with a  $P$ -value threshold of  $1 \times 10^{-9}$ , and with other settings as default parameters. Peak finding for each medulloblastoma was performed separately and as a control background for each H3K27ac ChIP-seq sample, its matched genomic DNA was used. The SPOT statistic<sup>44</sup>, a measure of read fraction found in enriched regions developed by the ENCODE consortium, was used to quantify H3K27ac enrichment quality. Primary medulloblastoma data sets had a median SPOT score of 0.62 which was equivalent to cell line data and on par with primary human data generated in the Epigenome ROADMAP. Afterwards, H3K27ac peaks were merged into a single coordinate file. Peaks which can not be identified in at least two primary medulloblastomas and contained completely within the region surrounding  $\pm 1$  kb TSS were excluded from any further analysis. This resulted in final combined and filtered peak set ( $n = 78516$ ). H3K27ac enrichments were calculated on the final peak set using the following formula:  $\log_2(((\text{Cnt}_{\text{ChIP}}/\text{LSize}_{\text{ChIP}} * \min(\text{LSize}_{\text{ChIP}}, \text{LSize}_{\text{ctrl}})) + \text{pscnt}) / ((\text{Cnt}_{\text{ctrl}}/\text{LSize}_{\text{ctrl}} * \min(\text{LSize}_{\text{ChIP}}, \text{LSize}_{\text{ctrl}})) + \text{pscnt}))$ , where  $\text{Cnt}_{\text{ChIP}}$  denotes the total number of reads mapping to the enhancer coordinate in ChIP sample,  $\text{LSize}_{\text{ChIP}}$  is the total library size for the ChIP sample,  $\text{Cnt}_{\text{ctrl}}$  is the total number of reads mapping to the enhancer coordinate in the control genomic DNA,  $\text{LSize}_{\text{ctrl}}$  is the total library size for the control sample, and  $\text{pscnt}$  is a constant number ( $\text{pscnt} = 8$ ), which was used to stabilize enrichments based on low read counts. (Peaks showing statistically significant differential H3K27ac enrichment across medulloblastoma subgroups were determined using ANOVA and the ones with  $\text{FDR} < 0.01$  were preserved after multiple testing correction. From the resulting peak-set, peaks having 1.5 ( $\log_2$ ) fold change difference across any medulloblastoma subgroup comparison were called as “subgroup specific” enhancers ( $n = 20,406$ ). Peaks that do not fulfil these criteria were referred as “common” enhancers ( $n = 58,110$ ). Subgroup-specific enhancers were further clustered using  $k$  means, with  $k = 6$  into 6 groups as “SHH”, “WNT”, “Group4”, “WNT-SHH”, “Group3-Group4”, and “Group3” (Fig. 2).

**Coverage of medulloblastoma enhancers in the genome.** Genome was classified into regions as exon, intron, intergenic and promoter (region surrounding  $\pm 1$  kb transcriptional start sites) by following the hierarchy promoter > exon > intron > intergenic. Then, medulloblastoma enhancers were intersected with these defined elements and fraction covered by each element was calculated.

**Enhancer saturation analysis.** To better understand whether our enhancer profiling adequately captured the primary medulloblastoma enhancer landscape, we performed a saturation analysis. We measured the total number of discreet regions and the fraction of novel regions gained by increasing sample number. This was performed across 1,000 permutations of the 28 medulloblastoma samples to establish 95% confidence intervals (Extended Data Fig. 1d).

**Comparison of H3K27ac with BRD4 occupancy and DNA methylation at enhancers.** Enrichment values for H3K27ac at enhancers were calculated as the ratio between library size normalized read counts for H3K27ac ChIP and its sample matched genomic DNA control. The formula used for the enrichment calculation is as follows:  $\log_2(((\text{Cnt}_{\text{ChIP}}/\text{LSize}_{\text{ChIP}} * \min(\text{LSize}_{\text{ChIP}}, \text{LSize}_{\text{ctrl}})) + \text{pscnt}) / ((\text{Cnt}_{\text{ctrl}}/\text{LSize}_{\text{ctrl}} * \min(\text{LSize}_{\text{ChIP}}, \text{LSize}_{\text{ctrl}})) + \text{pscnt}))$ , where  $\text{Cnt}_{\text{ChIP}}$  denotes the total number of reads mapping to the enhancer coordinate in ChIP sample,  $\text{LSize}_{\text{ChIP}}$  is the total library size for the ChIP sample,  $\text{Cnt}_{\text{ctrl}}$  is the total number of reads mapping to the enhancer coordinate in the control genomic DNA,  $\text{LSize}_{\text{ctrl}}$  is the total library size for the control sample, and  $\text{pscnt}$  is a constant number ( $\text{pscnt} = 8$ ), which was used

to stabilize enrichments based on low read counts. To compare BRD4 enrichment with H3K27ac enrichment at the enhancers, BRD4 enrichments were calculated in the same way as H3K27ac enrichments. DNA methylation values at enhancers were determined by calculating the average DNA methylation of all medulloblastoma samples where DNA methylation data are available<sup>6</sup>.

**Comparison of H3K27ac occupancy with H3K4me1, H3K27me3 and BRD4 occupancy.** We generated ChIP-seq data for H3K4me1 and H3K27me3 for only three Group 3 medulloblastomas (MB-1M21, MB-4M23, and MB-4M26). Therefore, comparison of H3K27ac occupancy with H3K4me1, H3K27me3 and BRD4 (Extended Data Fig. 1f) was done using the data from only these three Group 3 samples. To analyse the occupancy of the marks at H3K27me3 enriched regions, we called H3K27me3 peaks using MACS. ChIP-seq reads covering each base pair either in the region  $\pm 5$  kb around Group 3-specific enhancer midpoints (Extended Data Fig. 1f top panel) or in the region  $\pm 5$  kb around H3K27me3 peak midpoints (Extended Data Fig. 1f bottom panel) were quantified. Read coverage was averaged in 100-bp windows along the regions and the values were scaled to arrange between 0–1. Resulting values were represented as heat maps.

**Comparison of H3K27ac peak calling using whole genome sequencing or whole cell extract backgrounds.** We repeated H3K27ac peak finding (running MACS with a  $P$ -value threshold of  $1 \times 10^{-9}$ , and with other settings as default parameters) for the two medulloblastomas (MB12 and MB200) using their input chromatin as the backgrounds instead of using their matched whole genome sequencing. Resulting set of peaks identified using whole chromatin extract were compared to the ones identified using whole genome sequencing in scatter plots in Extended Data Fig. 1c. **Comparison of medulloblastoma H3K27ac enhancers with published H3K27ac data.** ENCODE<sup>8</sup> H3K27ac peaks were downloaded from [http://ftp.ebi.ac.uk/pub/databases/ensembl/encode/integration\\_data\\_jan2011/byDataType/peaks/jan2011/histone\\_macs/optimal/hub/](http://ftp.ebi.ac.uk/pub/databases/ensembl/encode/integration_data_jan2011/byDataType/peaks/jan2011/histone_macs/optimal/hub/) and all peaks were merged into a single coordinate file. Regarding ROADMAP data<sup>45,46</sup>, all available H3K27ac alignment files were downloaded and peak finding on individual samples was performed using MACS<sup>12</sup>. All ROADMAP H3K27ac peaks were as well merged into a single coordinate file. Resulting peaks from both ENCODE and ROADMAP were intersected with medulloblastoma H3K27ac peaks (with a minimum 50% overlap criteria; Fig. 1e, f).

**Comparison of medulloblastoma H3K27ac enhancers with CNV data.** To determine the overlap of enhancer loci with CNVs, medulloblastoma enhancer loci were intersected with focal amplifications and deletions obtained from<sup>4</sup>. To determine the statistical significance of the overlap, we performed 10,000 random simulations whereby CNV locations were randomly permuted across the genome without overlap using the bedtools shuffle utility (<http://bedtools.readthedocs.org>) and excluding regions found in the ENCODE<sup>8</sup> blacklist (<https://sites.google.com/site/anshulkundaje/projects/blacklists>). This distribution of random overlaps was used to calculate an empirical  $P$ -value of the observed overlap significance (Extended Data Fig. 1g).

**Quantification of gene expression and assignment of subgroup specific expression.** Expression values in RPKM were calculated using “qCount” function of Bioconductor package “quasR” (<http://www.bioconductor.org/packages/release/bioc/html/QuasR.html>). Genes showing differential gene expression across four medulloblastoma subgroups were determined using ANOVA ( $\text{FDR}$  less than 1%). Then, subgroup specific assignment of gene expression was done by performing a post-hoc test (using “glht” function of R package “multcomp”<sup>56</sup>).

**Identification of enhancer target genes.** Target gene identification of enhancers was performed as described<sup>47</sup>. For each enhancer, topology-associated domain (TAD)<sup>18</sup> which it belongs to was identified. Then, genes with transcriptional start sites falling into the same TAD were determined. We filtered nearby genes for protein coding status, as eRNAs and other enhancer associated ncRNAs are likely to emanate from enhancers and obfuscate distal target genes. Correlation tests (Spearman’s rank correlation coefficient) for H3K27ac enrichment of the enhancer and expression level of genes which are in the same TAD were performed. After repeating this procedure for each enhancer, all  $P$ -values obtained via correlation tests were combined and corrected for multiple testing globally using Bioconductor package “qvalue” (<http://www.bioconductor.org/packages/release/bioc/html/qvalue.html>). Correlations with a  $\text{FDR}$  less than 5% were preserved. For each enhancer, gene whose expression best correlates with the H3K27ac enrichment of the enhancer was selected as the potential target gene. For the cases where the difference between spearman correlation coefficients for the best and second best correlating genes were less than 0.1, the second best correlating gene was also selected as another potential target gene. Identification of enhancer target genes was performed for subgroup specific and common enhancers separately. After getting final gene lists for targets of subgroup specific and common enhancers, genes which are identified as targets both for subgroup specific and common enhancers were removed from common enhancer target gene list.



**Classification of enhancer targets according to enhancer regulation.** Genes regulated by differential enhancers were classified into categories depending on the number of differential enhancers they are targeted by (Fig. 2d). As mentioned in “identification of enhancer targets” part, to assign the enhancers to their targets with highest probability, in the final list of enhancer target genes, number of genes per enhancer was restricted to 2 genes having the highest correlation coefficient. However, to evaluate the number of genes targeted by each enhancer overall, enhancers were classified into categories depending on the number of genes they target by including all the genes targeted by enhancers (satisfying FDR < 0.05 criteria) (Fig. 2e).

**Overlap of target genes with regulatory information from literature.** Medulloblastoma signature genes were defined to be the genes regulated differentially in 4 medulloblastoma subgroups<sup>16</sup>. To be conservative on the signature genes, for each medulloblastoma subgroup, top 100 genes differentially regulated in the respective subgroups were included in the analysis. Resulting gene list were compared to the genes regulated by medulloblastoma subgroup specific enhancers and super-enhancers. Comparison to cancer genes was performed using the gene list provide in cancer gene census (<http://cancer.sanger.ac.uk/cancergenome/projects/census>). Target genes were overlapped with consensus TFs provided<sup>48</sup>. Inference whether the target genes we identified was druggable was done by intersecting target genes with the genes provided in the drug gene interaction database (<http://dgidb.genome.wustl.edu/>) by using “Expert curated” option in the source trust level category of the interactions. All information showing the overlap of target genes with gene lists from literature can be found in Supplementary Table 3.

**Pathway analysis.** Functional characterization of enhancer/gene assignments was conducted using the ClueGO plugin for cytoscape<sup>49</sup>. Subgroup-specific enhancer gene targets or SE-regulated TFs were queried against a compendium of gene sets from GO (Biological Process), KEGG, and REACTOME to identify processes/pathways that were significantly enriched in tested gene lists from our data set. Analyses were performed using the GO Term Fusion option in ClueGO and only processes/pathways with a *P*-value < 0.05 (right-sided hypergeometric test) following *P*-value correction (Bonferroni step down) were visualized. Manual trimming of ClueGO output was performed to remove processes/pathways affiliated with only a single gene set.

**Functional comparison of Group 3 and Group 4 enhancers.** To identify subgroup specific enhancers and their associated functional pathways, we performed a differential enhancer analysis<sup>50</sup> on Group 3 and Group 4 enhancers. We first took the union of the top 1,000 enhancer in Group 3 and Group 4 as defined by total H3K27ac signal (area under the curve). We next ranked all enhancer regions by the log<sub>2</sub> fold change in H3K27ac (Extended Data Fig. 3b). Differential enhancer target genes as previously defined were depicted under associated enhancers. Visual inspection revealed a number of TGF- $\beta$  pathway components associated with Group 3 specific enhancers. We visualized this by identifying all enhancer regulated TGF- $\beta$  pathway components (obtained from KEGG, REACTOME, and GO Biological Process databases) and depicting their specific regulation by Group 3, Group 4, or Group 3-4 differential enhancers (Extended Data Fig. 3c).

**Comparing enhancer acetylation at TGF- $\beta$  pathway components in ACVR2A amplified vs non-amplified Group 3 tumours.** We identified a focal amplification of the TGF- $\beta$  pathway receptor gene ACVR2A in the Group 3 medulloblastoma sample MB-4M23. Whole genome sequencing log<sub>2</sub> read depth ratio is plotted in Extended Data Fig. 3d. We hypothesized that in MB-4M23, amplification of ACVR2A leads to increased TGF- $\beta$  pathway activity, including the increased H3K27ac at enhancers regulating TGF- $\beta$  pathway components. We identified all Group 3 enhancers regulating TGF- $\beta$  pathway components and compared the median enhancer normalized H3K27ac signal in MB-4M23 vs all other Group 3 medulloblastomas. Extended Data Fig. 3e shows all enhancers ranked by their log<sub>2</sub> fold change in H3K27ac for MB-4M23 vs other Group 3 samples. The standard error of the mean was calculated for the fold change and is displayed as error bars in Extended Data Fig. 3e.

**Nucleosome free region (NFR) identification.** H3K27ac data for the samples within the same subgroup was combined. Nucleosome free regions per subgroup were identified by feeding these combine data sets to HOMER software (<http://homer.salk.edu/homer/ngs/index.html>) using “findPeaks” function with the option “-nfr”.

**Enrichment of TFs at subgroup-specific enhancers.** TF binding sites obtained from TRANSFAC<sup>51</sup> and detected at NFRs using FIMO<sup>52</sup> were overlapped with NFRs located within each class of differentially regulated enhancers. For each TF, contingency tables showing the number of NFRs overlapping and non-overlapping with the respective TF were constructed. Significance of enrichment of TFs in NFRs of differentially regulated enhancers was determined using Chi-squared test. Resulting *P*-values were corrected for multiple testing (FDR < 0.01). TF enrichments were calculated as the ratio between observed counts over expected counts. To represent TF enrichments as a heat map (Extended Data Fig. 6b), for each class of enhancers, 4–5 TFs showing the highest enrichments were selected.

**Linking subgroup-specific enhancers with TFs.** For each of differentially regulated enhancers in the classes of WNT, SHH, Group 3 and Group 4, NFRs belonging to each subgroup were overlapped with the respective subgroup-specific enhancers targeting at least one gene. Overlapping NFRs were intersected with TF binding sites having top 20th percentile enrichment scores in the respective subgroup-specific enhancers and differentially expressed in the same subgroup. For each TF, NFRs having the top 10th percentile number of binding sites were identified as sites occupied by the respective TF. Then, resulting NFRs were linked back to enhancers they are located, which enabled the linking of TFs having binding sites in the respective enhancers with the target genes of the enhancers. TF regulatory networks for each subgroup (Extended Data Fig. 7), where TFs represented as “sources” and enhancer target genes represented as “targets” were constructed using visualization platform Gephi (<http://gephi.github.io/>). To connect LMX1A, LHX2 and EOMES with their targets (Extended Data Fig. 9b), same strategy was applied by restricting the initial set of TFs to only those three.

**4C-seq data analysis.** Aligned 4C data was further processed, filtered and visualized using Bioconductor package “Basic4Cseq”<sup>53</sup>.

**Mapping typical enhancers and super-enhancers using H3K27ac enhancer definitions.** H3K27ac super-enhancers (SEs) and typical enhancers (TEs) in individual medulloblastoma samples were mapped using the ROSE2 software package described<sup>13,23</sup> and available at <https://github.com/BradnerLab/pipeline>. A 12.5 kb stitching window was used to connect proximal clusters of H3K27ac peaks into contiguous enhancer regions. These mappings identified on average ~600 SEs per sample.

**Clustering medulloblastoma samples by SE patterns.** Relationships between SE landscapes between samples were determined as in ref. 11. First, we defined the union of all regions considered to be an SE in any individual primary sample and in three Group 3 cell lines. Next H3K27ac signal was calculated at each region and median normalized for each sample. Samples were hierarchically clustered based on similarity of patterns of median normalized H3K27ac enhancer signal as determined using pairwise Pearson correlations.

**Mapping SEs and typical enhancers across medulloblastoma subgroups (subgroup enhancer mapping).** In order to map and quantify enhancer regions for each medulloblastoma subgroup, we first mapped all enhancers in each individual sample within the group. Across a group, we used the union of all enhancer regions within group samples as the landscape of enhancers. Within this landscape, enhancers were ranked by average H3K27ac signal (area under curve) and classified as SEs or TEs as previously described. This produced SE and TE meta enhancer landscapes for WNT, SHH, Group 3, and Group 4 medulloblastoma with between 558 and 1,110 SEs called per group (Fig. 3a). Locations for all SEs and TEs in each subgroup are provided in Supplementary Table 4.

**Quantifying enhancer signal variance across samples at meta enhancer regions.** To compare the dynamic range of SEs and TEs defined in each medulloblastoma subgroup, we quantified H3K27ac signal variance across samples. For SE and TE enhancer constituents (individual peaks of H3K27ac enrichment within broader enhancer domains) defined in each group, H3K27ac signal variance across samples as a fraction of the mean sample was calculated. The average H3K27ac signal variance across all SEs or TEs within a group is plotted in Extended Data Fig. 4f.

**Quantifying average H3K27ac signal across samples at subgroup SEs and typical enhancers.** We sought to examine trends in H3K27ac signal across medulloblastoma samples at regions defined as SEs or TEs in each group. First we mapped H3K27ac across all samples to enhancer constituents defined in each group. For each medulloblastoma sample, the average median normalized H3K27ac signal was plotted for SE and TE constituents respectively. For SEs and TEs defined in each group, the average sample H3K27ac signal is plotted with the mean and standard deviation shown as lines. This visualization enables a rapid assessment of H3K27ac variance within a group and of trends in H3K27ac signal for SEs and TEs defined in each group (Extended Data Fig. 4h). For instance, enhancer constituents in Group 3 SEs tend to have high signal in Group 4.

**Quantifying group ChIP-seq signal at subgroup SEs and typical enhancers within and between groups.** SEs have been shown to have higher H3K27ac and BRD4 signal density at constituents when compared to typical enhancers<sup>13,23</sup>. To determine if these trends were observed at medulloblastoma enhancers, we calculated H3K27ac and BRD4 ChIP-seq signal density across all samples at all regions defined as enhancers across groups (meta enhancers). In order to properly compare ChIP-seq signal density between SEs and TEs, for each enhancer constituent, we first determined if it was considered part of an SE in one or more groups, and if so, these groups defined the “active group context” for that particular enhancer constituent. Groups in which the enhancer constituent showed no evidence of enhancer activity (SE or TE) were considered the inactive group context. For enhancer constituents considered only part of a TE in one or more groups, groups in which the enhancer constituent was classified as a TE were considered the active

group context and all other groups were considered the inactive group context. For each SE or TE constituent, average H3K27ac or BRD4 signal density was calculated at all samples in the active group context or in the inactive group context. The distributions of H3K27ac or BRD4 signal for enhancer constituents classified by SE or TE status were plotted and the statistical significance of the difference in the mean was tested in the active or inactive group context using a Welch's two-tailed *t*-test (Extended Data Fig. 4g).

**Identifying group specific and conserved SEs.** We developed a method to identify SEs that were conserved across all medulloblastoma subgroups as well as SEs that showed highly group specific patterns of enhancer activity. We first took as the SE landscape all regions identified as SEs in the meta subgroup enhancer mapping. To account for sample-to-sample variability in H3K27ac ChIP-seq dynamic range, H3K27ac signal at enhancers in each medulloblastoma sample was rank transformed (Fig. 3b). As each medulloblastoma sample contained on average ~600 SEs, enhancer regions with an average rank of 600 or better in each subgroup were considered conserved. To identify enhancers with group specific patterns of activity, we calculated a "group rank Z-score" that compared average signal in one group to average signal in other groups. Here we considered whether enhancers might show group specific patterns for WNT, SHH, Group 3, Group 4, and as well for groupings of WNT/SHH, and Group 3/4. For each enhancer, this group rank Z-score was calculated for each group vs other combination. Enhancers with a group rank Z-score > 1 (that is, those whose mean rank within a group was > 1 standard deviation above the mean rank of all other samples) were considered group specific. To account for variability in enhancer ranks, only enhancers with a statistically significant difference in ranks (within group vs all other samples, Welch's two-tailed *t* test, *P*-value < 0.01) were considered. Supplementary Table 4 contains all SE regions identified in medulloblastoma subgroups and their corresponding max group rank Z-score, *P*-value, and classification.

**Mapping H3K27ac enrichment at the MYC gene desert.** To provide a developmental context for medulloblastoma MYC SEs, we mapped H3K27ac enrichment at the MYC locus. H3K27ac data was obtained from the Epigenome ROADMAP as in Fig. 1e. The 500 kb region flanking the MYC SE No. 2 was divided into 5 kb bins and each bin was tested for overlap with a H3K27ac peak in each ROADMAP sample. ROADMAP samples were hierarchically clustered by similarity of H3K27ac peak pattern at the MYC locus (Extended Data Fig. 5m). Overlap with MYC SE No. 2 was found in 4/77 ROADMAP samples.

**Calculating regulatory IN and OUT degree for all SE associated TFs.** Medulloblastoma core regulatory circuitry analysis was performed using the COLTRON (<https://pypi.python.org/pypi/coltron>) that calculated inward and outward degree regulation of SE-regulated TFs. To quantify the interaction network of TF regulation, we calculated the IN and OUT degree of all SE associated TFs. The 92 SE associated TFs were those defined as either proximal to an SE (within 50 kb) or the target of a differential SE enhancer element. For any given TF (TF<sub>i</sub>), the IN degree was defined as the number of TFs with an enriched binding motif at the proximal SE of TF<sub>i</sub> (Fig. 5a). The OUT degree was defined as the number of TF associated SEs containing an enriched binding site for TF<sub>i</sub>. Within any given SE, enriched TF binding sites were determined at putative nucleosome free regions (valleys) flanked by high levels of H3K27ac. Valleys were calculated using an algorithm adapted from ref. 35. In these regions, we searched for enriched TF binding sites using the FIMO<sup>52</sup> algorithm with TF position weight matrices defined in the TRANSFAC database<sup>51</sup>. An FDR cutoff of 0.01 was used to identify enriched TF binding sites. Using this approach, we calculated IN and OUT degree for all SE associated TFs within the meta H3K27ac landscape (average of all samples) of each medulloblastoma subgroup. This approach resulted in an IN and OUT degree estimate for each SE associated TF in each medulloblastoma subgroup (Extended Data Fig. 8a–d).

**Identifying TF binding motifs for LMX1A, LHX2, and HLX.** We sought to identify TF binding motifs for each TF in each subgroup. For each TF, we defined binding regions as the ±1,000 bp flanking the enriched region summit (as defined using MACS 1.4.2 with a *P*-value cutoff of  $1 \times 10^{-9}$ ). We took the union of all regions bound in a given subgroup (for example, HLX bound regions in Group 3 samples) that overlapped an enhancer in that subgroup and did not overlap any ENCODE<sup>8</sup> blacklist regions. We next took the top 10,000 discreet regions as ranked by average TF ChIP-seq signal and used the ±100 bp region flanking the region centre as the input for *de novo* motif finding. *De novo* motif finding was performed using the MEME<sup>54</sup> suite using a 1st order background model and searching for motifs between 6 and 30 bp in length. The top motif for each TF is displayed as a position weight matrix in Extended Data Fig. 8i–l.

**Visualizing TF regulatory networks.** To visualize SE associated TF interactions in each subgroup, we ranked all SE associated TF by TOTAL degree (IN + OUT).

We visualized the top 50% of SE associated TFs in each subgroup as a network diagram with each node representing a SE associated TF, and with nodes coloured and ordered by increasing TOTAL degree (Extended Data Fig. 8e–h). Interactions between SE associated TF nodes were defined as a TF motif identified in the SE of a TF and are depicted as edges. For Group 3 and Group 4, edges validated by the presence of a TF ChIP-seq peak are coloured.

**Clustering TFs by regulatory degree to identify and infer subgroup specific regulatory circuitry.** To identify SE associated TFs with similar regulatory patterns likely to influence subgroup identity, we first normalized the TOTAL degree for each SE associated TF in each subgroup from 0 to 1. We then calculated the normalized TOTAL degree for each SE associated TF in each subgroup. We filtered out all TFs with a max TOTAL degree across medulloblastomas of less than 0.7.

We next clustered all remaining TFs by their TOTAL degree pattern. Hierarchical clustering was performed using a Euclidian distance metric and the resulting clustergram tree was cut at a distance of 0.5 to produce 26 individual clusters. Of these 26 clusters, 12 showed a median TOTAL degree > 0.7 in 1, 2, or all 4 subgroups. Clusters with > 0.7 TOTAL degree in 3 subgroups were omitted for simplicity. TOTAL degree patterns of TFs in these 12 clusters are shown in Extended Data Fig. 9a. This filtering produced a list of 102 SE associated TFs, of which 71 had predicted interactions with one another. These 71 TFs fall into either conserved, subgroup specific, or dual subgroup clusters and together they comprise the inferred core regulatory circuitry of medulloblastoma subgroups. As in Extended Data Fig. 8e–h, regulatory interactions between these core regulatory circuitry TFs are depicted in Extended Data Fig. 9a with Group 3 and Group 4 validated edges coloured. A subset of this larger network containing the TFs HLX, LHX2, EOMES, and LMX1A is depicted in Fig. 5c with ChIP-seq validated edges drawn as solid lines and motif prediction edges drawn in dotted lines.

**Quantifying protein–protein interactions of co-regulating SEs.** We used the STRING interaction database<sup>55</sup> to quantify protein–protein interaction frequencies of SE associated TFs with similar regulatory patterns. TF pairs were considered co-regulatory if they shared 50% of the same OUT degree edges. Interaction frequencies for co-regulatory pairs were compared to those from 10,000 randomly assigned pairs of TFs expressed in that subgroup (Extended Data Fig. 8o).

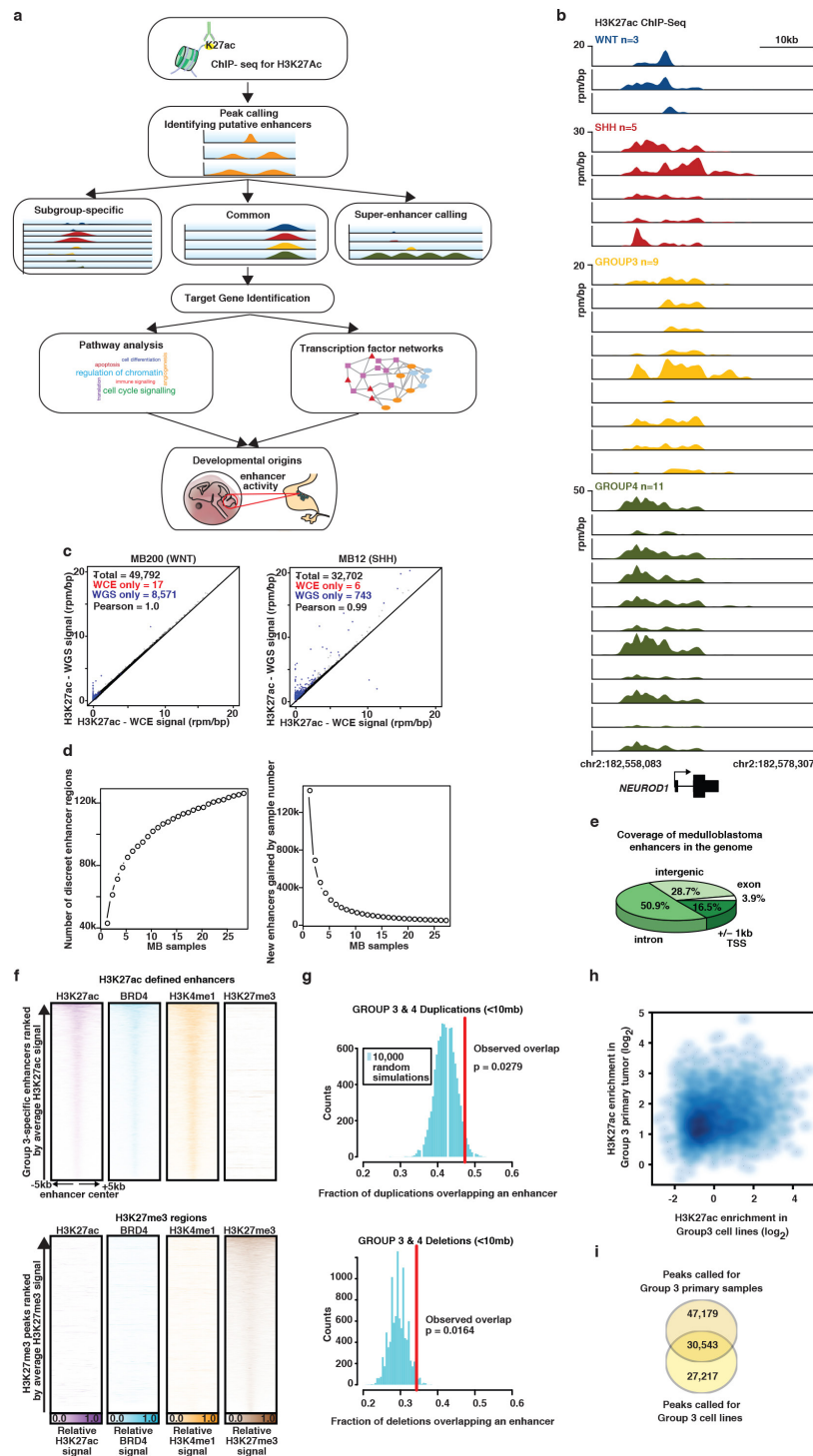
**Integration of TF ChIP-seq occupancy into enhancer landscape and TF regulatory network.** To determine the fraction of motif predicted edges with evidence of actual TF ChIP-seq binding, we first took all predicted edges for HLX, LHX2, and LMX1A interacting SE associated with other TFs in Group 3 and Group 4. We validated all edges that contained a ChIP-seq peak within the same enhancer as the predicted TF motif. The fraction of validated edges for each TF in each subgroup is shown in Extended Data Fig. 8g, h, m.

**Quantification of TF binding at Group 3 and 4 enhancers.** To determine how Group 3 and Group 4 TF ChIP-seq levels varied at Group 3 and Group 4 specific enhancers, we quantified TF ChIP-seq signal at Group 3 and Group 4 enhancers. We first took the union of the top 1,000 enhancer regions as defined by H3K27ac signal in Group 3 and Group 4 (as in Extended Data Fig. 3b). We identified as Group 3 and Group 4 specific enhancer regions with a > 1.0 log<sub>2</sub> absolute fold change between Group 3 and Group 4. We identified as conserved enhancer regions with a < 0.05 log<sub>2</sub> absolute fold change between Group 3 and Group 4. We next identified all enhancer regions bound by LHX2 and HLX in Group 3 (G3 HLX and LHX2) or by LHX2 and LMX1A in Group 4 (G4 LMX1A and LHX2). TF ChIP-seq occupancy in units of average area under the curve (AUC) were quantified at TF bound regions overlapping Group 3 specific, Group 4 specific, and conserved enhancer region (Extended Data Fig. 8n). Statistical differences in the means of the distributions of TF ChIP-seq signal at different enhancer populations was determined using a Welch's two tailed *t*-test (Extended Data Fig. 8n).

**Quantifying Group 4 TF gene expression changes in Dreher RL.** To identify genes transcriptionally regulated by *Lmx1a* in the developing cerebellum, we isolated cerebellar uRL from WT and *Lmx1a*<sup>−/−</sup> embryos by laser capture microdissection. uRL was isolated from WT (*n* = 3) and *Lmx1a*<sup>−/−</sup> (*n* = 3) embryos (~3,000 cells per embryo) at e13.5, just before abnormal RL regression in *Lmx1a*<sup>−/−</sup> embryos. RNA was extracted using PicoPure RNA Isolation Kit (Arcturus) and hybridized to Illumina MouseRef8 v2 Expression BeadChips at the Johns Hopkins Array Core Facility. Next we identified all human TF genes with unambiguous mouse homologues that were detectably expressed in the WT mouse cerebellum (cut off of 100 arbitrary units). We subsequently quantified median normalized expression in WT or *Lmx1a*<sup>−/−</sup> samples and calculated the log<sub>2</sub> fold-change for all TFs. We ranked the expression fold-change of all SE-associated TFs in medulloblastoma and plotted their log<sub>2</sub> fold change in *Lmx1a*<sup>−/−</sup> vs WT (Fig. 6d). SE-associated TFs present in the Group 4 TF network (Extended Data Fig. 8h) were coloured in green.

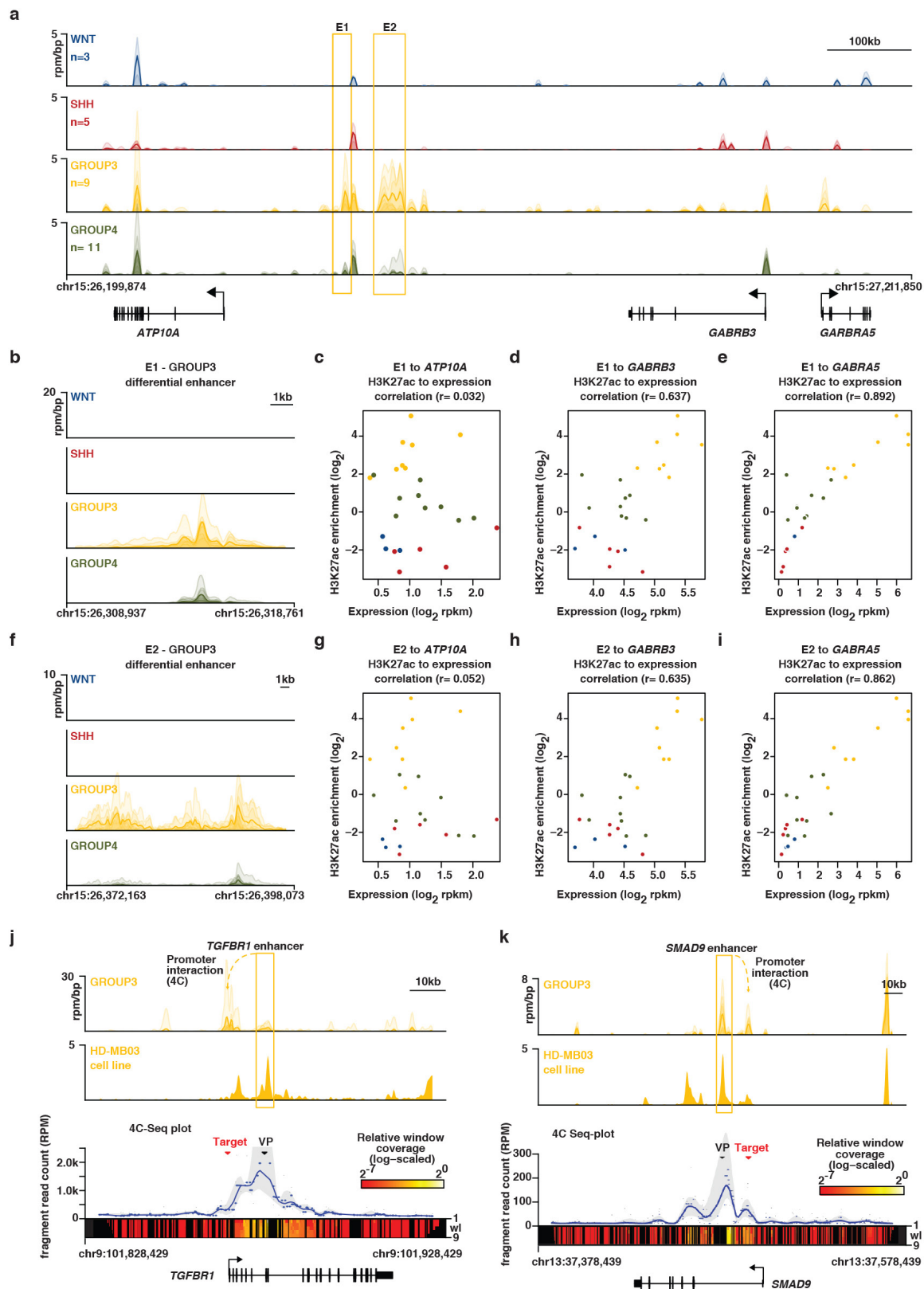
35. Ramsey, S. A. *et al.* Genome-wide histone acetylation data improve prediction of mammalian transcription factor binding sites. *Bioinformatics* **26**, 2071–2075 (2010).
36. Ni, T. T. *et al.* Conditional control of gene function by an invertible gene trap in zebrafish. *Proc. Natl Acad. Sci. USA* **109**, 15389–15394 (2012).
37. Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488**, 43–48 (2012).
38. Haldipur, P. *et al.* Expression of Sonic hedgehog during cell proliferation in the human cerebellum. *Stem Cells Dev.* **21**, 1059–1068 (2012).
39. Hovestadt, V. *et al.* Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol.* **125**, 913–916 (2013).
40. Borodina, T. & Adjaye, J. & Sultan, M. A strand-specific library preparation protocol for RNA sequencing. *Methods Enzymol.* **500**, 79–98 (2011).
41. Sultan, M. *et al.* A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods. *Biochem. Biophys. Res. Commun.* **422**, 643–646 (2012).
42. van de Werken, H. J. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89–112 (2012).
43. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
44. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genet.* **43**, 264–268 (2011).
45. Romanoski, C. E., Glass, C. K., Stunnenberg, H. G., Wilson, L. & Almouzni, G. Epigenomics: Roadmap for regulation. *Nature* **518**, 314–316 (2015).
46. Skipper, M. *et al.* Presenting the epigenome roadmap. *Nature* **518**, 313 (2015).
47. Waszak, S. M. *et al.* Population variation and genetic control of modular chromatin architecture in humans. *Cell* **162**, 1039–1050 (2015).
48. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nature Rev. Genet.* **10**, 252–263 (2009).
49. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
50. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363 (2008).
51. Matys, V. *et al.* TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–D110 (2006).
52. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
53. Walter, C., Schuetzmann, D., Rosenbauer, F. & Dugas, M. Basic4Cseq: an R/Bioconductor package for analyzing 4C-seq data. *Bioinformatics* **30**, 3268–3269 (2014).
54. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
55. Franceschini, A. *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–D815 (2013).





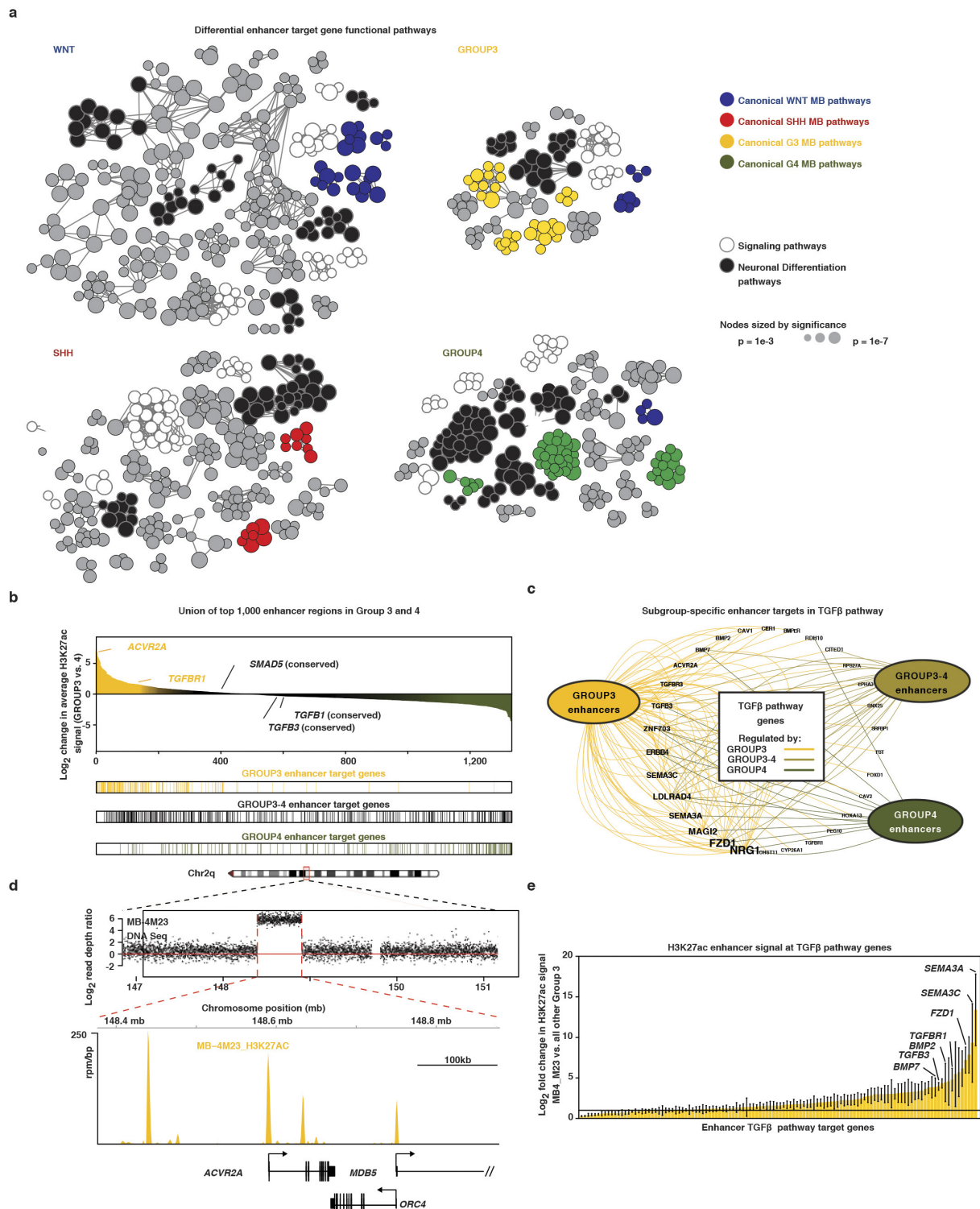
**Extended Data Figure 1 | Enhancer landscape of primary medulloblastoma.** This figure accompanies Fig. 1. **a**, Experimental workflow for studying enhancers and super-enhancers in primary medulloblastomas. **b**, H3K27ac ChIP-seq data showing a highly active enhancer at the *NEUROD1* locus across all 28 primary medulloblastoma samples from our series. **c**, Scatter plots showing Pearson correlation of H3K27ac peaks called using either sample-matched WGS or whole-cell extract (WCE) sequences as background for two samples from our series. **d**, Saturation analysis showing the number of discrete enhancer regions identified as a function of increasing sample number (top), or the fraction of newly gained discrete enhancer regions as a function of increasing sample number (bottom). Error bars represent 95% confidence intervals obtained from 1,000 permutations of sample order. **e**, Pie chart showing the genomic distribution of enhancer elements in medulloblastoma.

**f**, Heat maps of ChIP-seq data showing the scaled read densities for H3K27ac, BRD4, H3K4me1, and H3K27me3 in regions located  $\pm 5$  kb from Group 3-specific H3K27ac (top panel) and H3K27me3 peak midpoints (bottom panel). **g**, Histograms showing the fractional overlap of enhancers with focal amplifications (top) or focal deletions (bottom) in Group 3 and Group 4 medulloblastoma samples. The blue distributions represent expected fractional overlap generated from 10,000 random simulations. The red line depicts the actual observed fractional overlap with empirical *P*-value noted. **h**, Scatter plot correlating average H3K27ac enrichment in Group 3 cell lines with average H3K27ac enrichment in Group 3 primary medulloblastomas. Enrichments are calculated for peaks called in primary Group 3 samples. **i**, Venn diagram showing the overlap between H3K27ac peaks called for primary Group 3 medulloblastomas and Group 3 medulloblastoma cell lines.



**Extended Data Figure 2 | Enhancer/gene assignments in medulloblastoma.** This figure accompanies Fig. 2. **a**, Meta H3K27ac ChIP-seq tracks of the Group 3-specific enhancers (E1 and E2) in the TAD containing *ATP10A*, *GABRB3*, and *GABRA5*. **b**, Zoom in meta H3K27ac ChIP-seq tracks of enhancer E1 from **a**. **c–e**, Scatter plots correlating sample-matched gene expression (log<sub>2</sub> RPKM, x-axis) of *ATP10A* (**c**), *GABRB3* (**d**), and *GABRA5* (**e**) with H3K27ac enrichment (log<sub>2</sub>; y-axis)

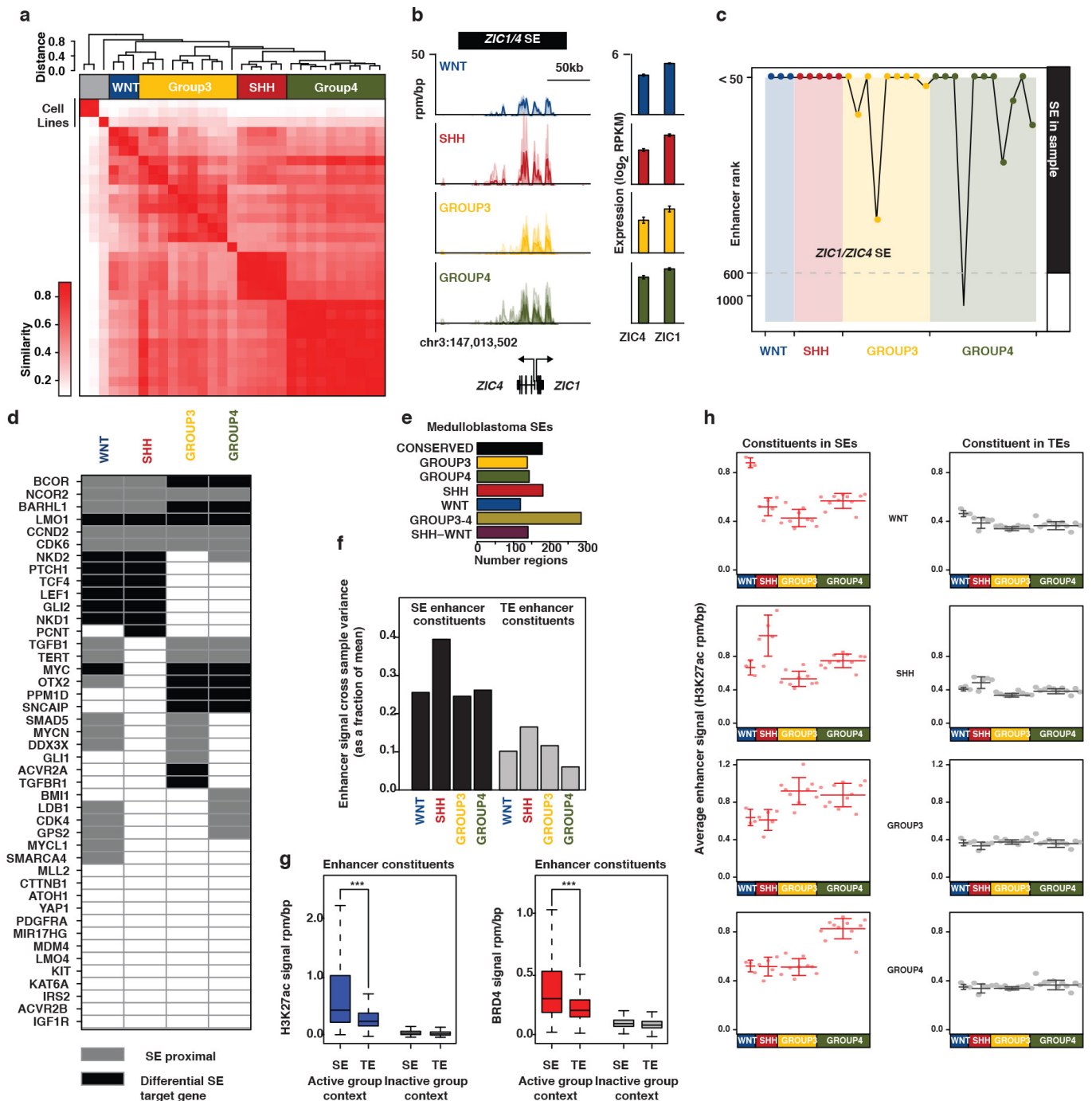
for the Group 3-specific enhancer shown in **b**, **f**, Zoom in meta H3K27ac ChIP-seq tracks of enhancers E2 from **a**. **g–i**, Scatter plots correlating sample-matched gene expression (log<sub>2</sub> RPKM, x-axis) of *ATP10A* (**g**), *GABRB3* (**h**), and *GABRA5* (**i**) with H3K27ac enrichment (log<sub>2</sub>; y-axis) for the Group 3-specific enhancer shown in **f**. **j**, **k**, 4C-seq validation of *TGFBR1* (**j**) and *SMAD9* (**k**) enhancer/promoter interactions in a Group 3 cell line (HD-MB03).



**Extended Data Figure 3 | Enhancer-driven TGF-β activity in Group 3 medulloblastoma.** This figure accompanies Fig. 2. **a**, Functional annotation of target genes assigned to subgroup-specific enhancers based on their significant overlap with gene sets annotated in Gene Ontology (GO Biological Process) and pathway databases (KEGG, Reactome). **b**, Waterfall plot discriminating the top 1,000 Group 3 and Group 4 subgroup-specific enhancers as defined by total H3K27ac signal. The distribution of assigned targets in Group 3, Group 4, and shared Group 3-4 targets are shown below the waterfall. **c**, Convergence of Group 3-specific enhancers on TGF-β pathway genes. Subgroup-specific enhancers are summarized as nodes according to their respective medulloblastoma enhancer class—Group 3, Group 4, and shared Group 3/Group 4—with

edges representing individual enhancer/TGF-β pathway gene assignments. **d**, Amplification of the TGF-β type II receptor, *ACVR2A*, in a Group 3 medulloblastoma from the ChIP-seq cohort (MB-4M23). Log<sub>2</sub> read depth data (tumour versus matched germline) derived from WGS data for this case is shown (upper panel). Highly active H3K27ac enhancer peaks overlapping the amplified *ACVR2A* locus are shown for the same case (lower panel). **e**, Bar plot showing the difference in H3K27ac enhancer signal between MB-4M23 (*ACVR2A*-amplified Group 3 sample) and all other Group 3 samples. Bar plot shows H3K27ac log<sub>2</sub> fold change at all enhancers regulating TGF-β component genes. Enhancers are ranked by increasing change in H3K27ac. Error bars represent standard error of the mean fold change.

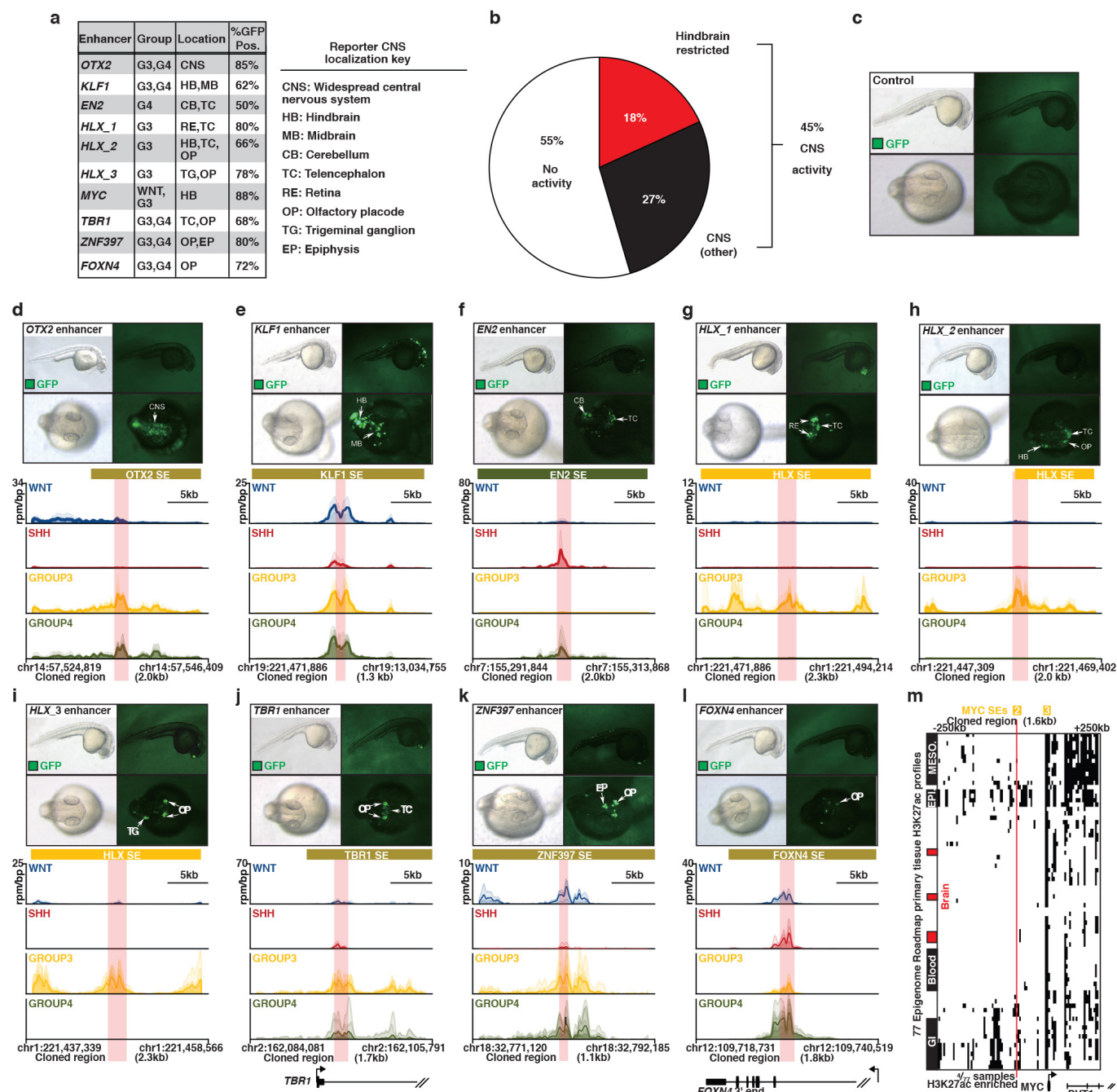




**Extended Data Figure 4 | Features of medulloblastoma super-enhancers.**

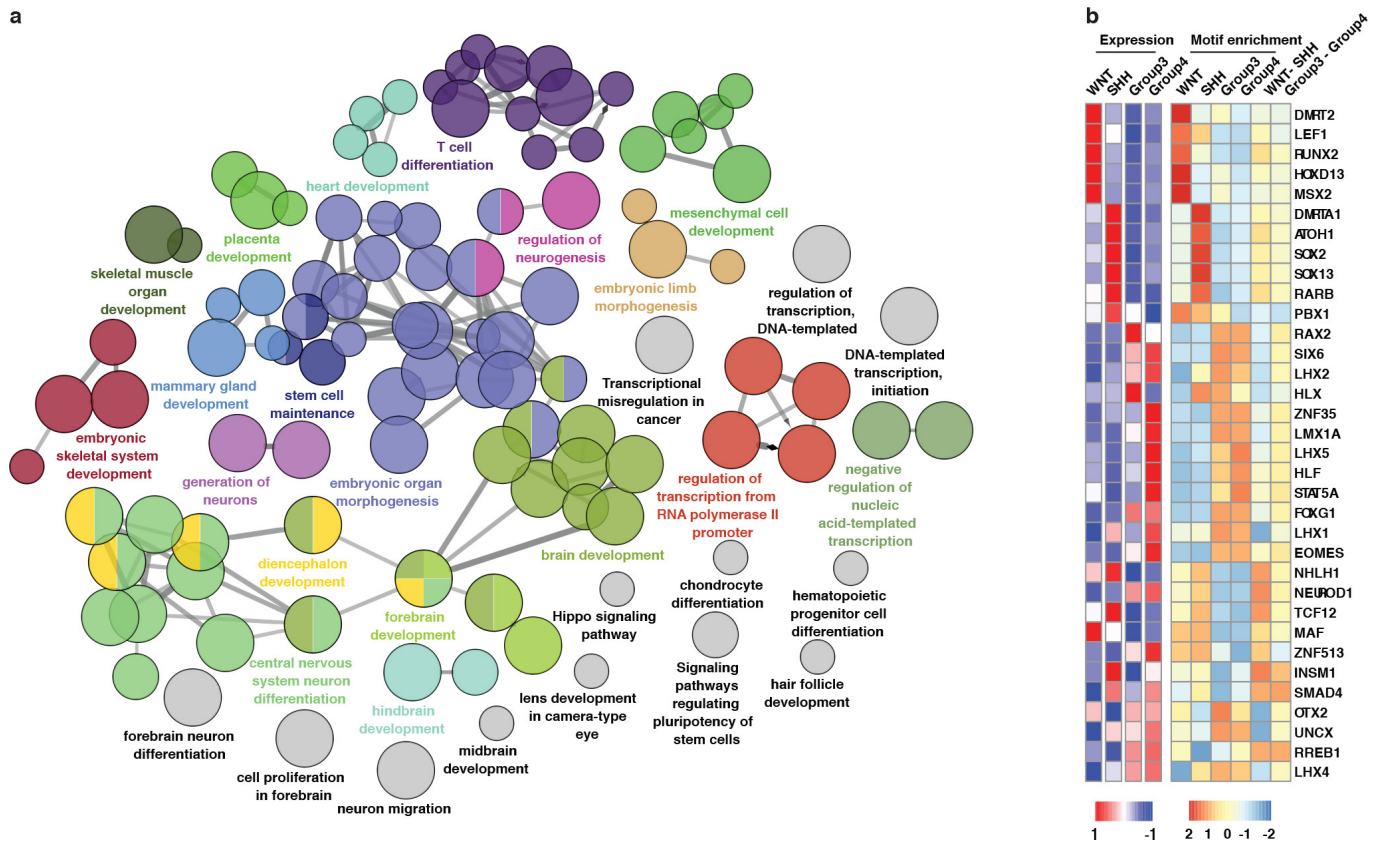
This figure accompanies Fig. 3. **a**, Unsupervised hierarchical clustering of primary medulloblastomas and cell lines using H3K27ac signal calculated at all SEs identified in each individual sample. **b**, Meta tracks of H3K27ac ChIP-seq signal for the *ZIC1/ZIC4* SE locus. Expression (mean RPKM) for both *ZIC4* (left) and *ZIC1* (right) is displayed as bar graphs to the right of each H3K27ac track with error bars representing s.d. of the mean ( $n = 140$  samples). **c**, Line plot showing the enhancer rank for the *ZIC1/ZIC4* SE locus across all samples according to subgroup. **d**, Heat map showing the SE association of known medulloblastoma driver genes and chromatin modifiers. Genes with called differential SEs are shaded black, whereas genes with proximal SEs (within 100 kb of TSS) are shaded grey, according to their respective subgroup. **e**, Bar plot showing the number of SE regions assigned to individual enhancer classes in medulloblastoma. **f**, Bar plot

of enhancer signal cross sample variance (y-axis) displayed as a fraction of the mean for SE enhancer constituents (left, black) or TE enhancer constituents (right, grey) identified in each medulloblastoma subgroup. **g**, Box plots of H3K27ac (left, blue) or BRD4 (right, red) enhancer signal at SEs or typical enhancers (TE) in their active group-specific context or in their inactive group context (for example, for SEs or TEs present in Group 3, active group context includes all Group 3 samples and inactive group context includes all other samples). Differences in the means of the distributions is quantified by a Welch's two-tailed  $t$ -test ( $***P < 1 \times 10^{-9}$ ). **h**, Dot plots of average H3K27ac enhancer signal in the constituents of SEs (left) or TEs (right) for enhancer constituents identified in WNT, SHH, Group 3, or Group 4 samples, respectively. Error bars represent standard deviation of the mean across all samples in a subgroup.



**Extended Data Figure 5 | *In vivo* validation of Group 3 and Group 4 medulloblastoma super-enhancers.** This figure accompanies Fig. 4. **a**, Summary of zebrafish reporter assays. **b**, Pie chart showing the fraction of all tested medulloblastoma enhancer regions that demonstrate any central nervous system localized reporter activity. **c–l**, Representative bright-field and fluorescence images of embryos (1 dpf) injected with individual enhancer-containing Tol2 vectors. Lateral views (60 $\times$ ) show GFP reporter expression in the whole body and dorsal views show GFP expression in the central nervous system (120 $\times$ ). White arrows indicate the locations of GFP signal. CNS, central nervous system; HB, hindbrain; MB, midbrain; CB, cerebellum; TC, telencephalon; RE, retina; OP, olfactory placode; TG, trigeminal ganglion. For each tested enhancer,

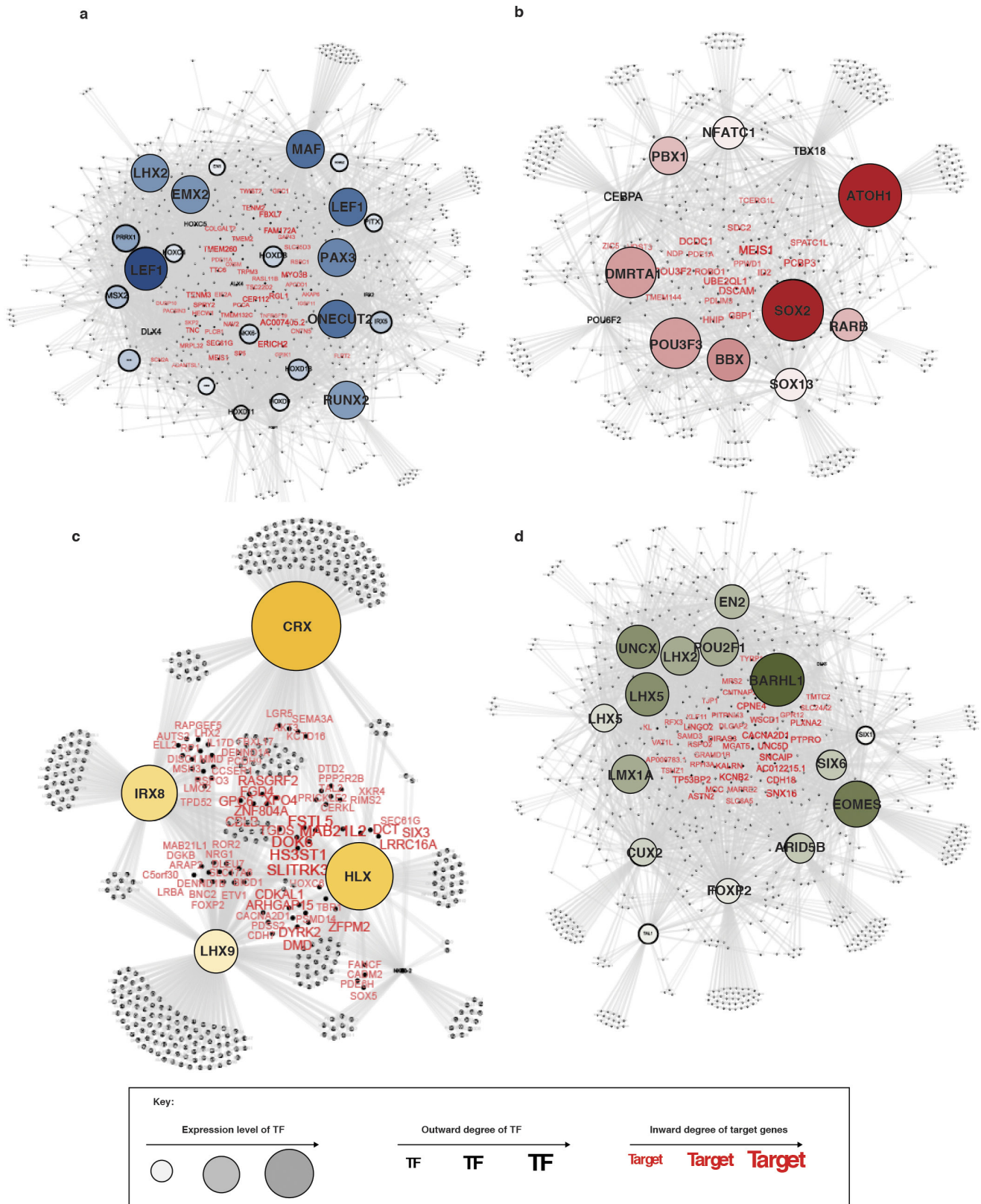
meta tracks of H3K27ac ChIP-seq signal across medulloblastoma subgroups for the cloned regulatory element are shown. **m**, Heat map showing H3K27ac enrichment at the  $\pm 250$  kb region flanking the medulloblastoma MYC SE described in Fig. 4 (SE #2; panels f, h–j) across 77 Epigenome Roadmap tissues. Each row represents a single tissue. Each column represents a region of the MYC gene desert locus. Black shaded regions indicate the presence of H3K27ac enrichment. The samples are ordered by similarity of H3K27ac enrichment pattern. Notable clusters of mesoderm (MESO.), epithelial (EPI.), blood, brain, or GI lineage derived samples are noted. The cloned enhancer reporter region described in Fig. 4 (panels f, h–j) is depicted as a vertical line and shows overlap with only 4/77 H3K27ac Epigenome Roadmap samples.



**Extended Data Figure 6 | Pathways regulated by super-enhancer associated transcription factors in medulloblastoma.** This figure accompanies Fig. 5. **a**, Functional pathways regulated by SE-associated TFs in medulloblastoma. **b**, Heat map of select subgroup-specific TFs

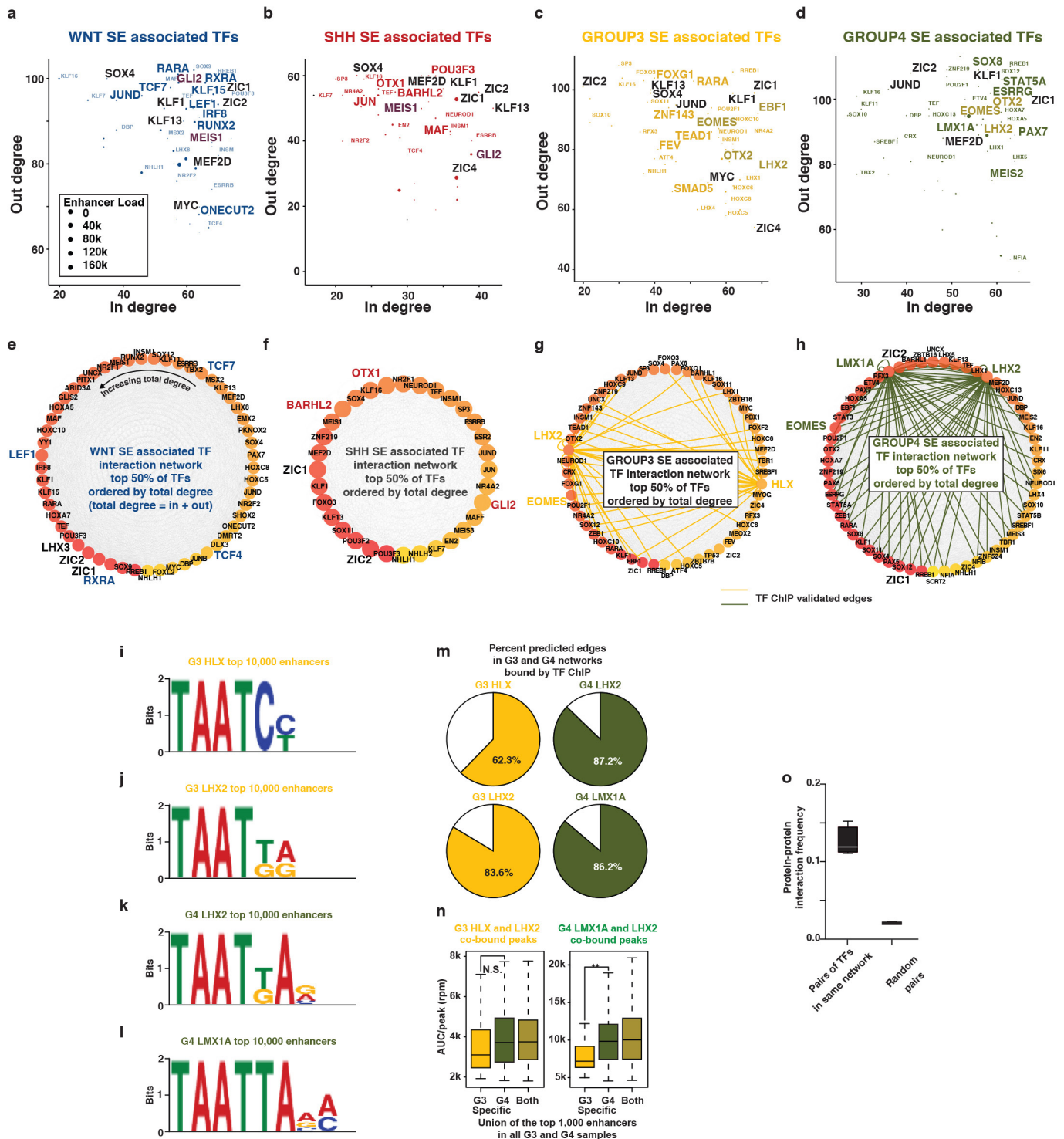
showing their expression (left columns) and enhancer motif enrichment (right columns). Enhancer motif enrichment was calculated at differential enhancer elements in the respective enhancer classes.





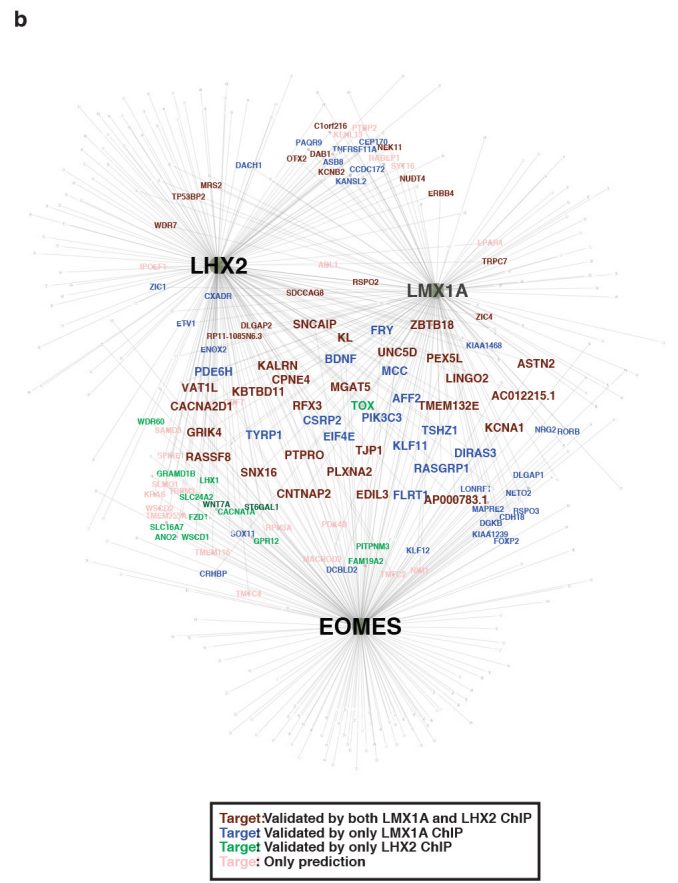
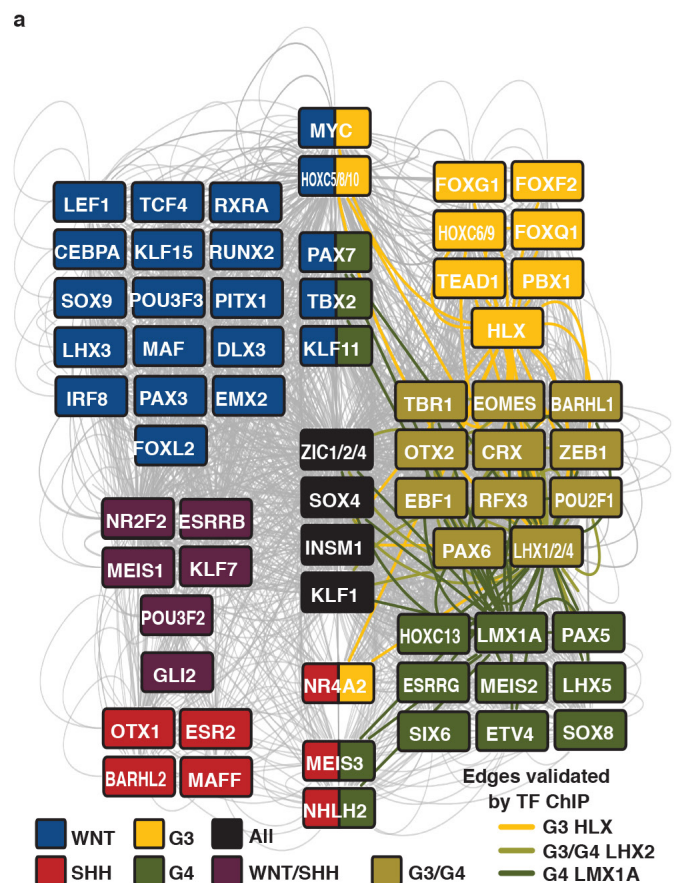
**Extended Data Figure 7 | Medulloblastoma subgroup-specific transcription factors and their associated target genes.** This figure accompanies Fig. 5. **a–d**, Network of subgroup-specific TFs and their predicted target genes for WNT (**a**), SHH (**b**), Group 3 (**c**) and Group 4 (**d**) subgroups. Nodes represent subgroup-specific TFs. In each subgroup,

node size is scaled and shaded according to the expression level of the TF and node font is scaled and shaded according to the number of inferred target genes (that is, OUT degree). TF target genes are shown in red font scaled according to the number of TFs predicted to target that gene (that is, IN degree).



**Extended Data Figure 8 | Super-enhancers define medulloblastoma regulatory circuitry.** This figure accompanies Fig. 5. **a–d**, Scatter plots of IN (x-axis) and OUT (y-axis) regulatory degree for SE-associated TFs in each medulloblastoma subgroup. **e–h**, TF interaction networks for each medulloblastoma subgroup. Nodes represent the top 50% of SE-associated TFs in each subgroup as ranked by total degree (counter clockwise). Each node is coloured by total degree and predicted binding interactions with other TF SEs are shown as edges. For Group 3 and Group 4 networks, edges validated by TF ChIP-seq binding are coloured. **i–l**, Position weight matrices showing the top statistically enriched motif identified for each transcription factor at the top 10,000 bound enhancers in each subgroup.

**m**, Pie charts showing the fraction of predicted edges in each Group 3 and Group 4 TF networks that are validated by the presence of the respective TF ChIP-seq binding at the enhancer. **n**, Medulloblastoma subgroup distribution of shared, co-bound peaks for master regulatory TFs analysed by ChIP-seq. TF binding is quantified as area under curve per peak (AUC/peak) in units of rpm. Differences in the means of the distributions is quantified by a Welch's two-tailed *t*-test (N.S.  $P > 0.1$ ,  $**P < 1 \times 10^{-6}$ ). **o**, Box plot of protein-protein interaction frequency (y-axis) calculated from STRING database for pairs of SE-associated TFs showing patterns of subgroup-specific SE co-regulation (left) or randomized pairs (right).



**Extended Data Figure 9 | LMX1A, EOMES, and LHX2 are master transcriptional regulators of Group 4 medulloblastoma.** This figure accompanies Fig. 5. **a**, Subgroup-specific regulatory circuitry. Nodes are TFs associated with an SE in a subgroup-specific context. Edges indicate co-regulating TFs as defined by enrichment of TF binding motifs in respective regulatory regions. Edges validated by TF ChIP-seq are coloured according to their respective subgroup association. **b**, Network involving

LHX2, LMX1A, and EOMES TFs and target genes inferred based on the presence of the respective TF motifs in Group 4-specific enhancers. Target genes are coloured according to their validation status based on LMX1A and LHX2 ChIP-seq, with genes arranged in the centre of the network inferred to be targeted by all three master TFs. For visualization purposes, these common targets are displayed with a larger font size compared to the genes in the surrounding network.



# A homogeneous nucleus for comet 67P/Churyumov–Gerasimenko from its gravity field

M. Pätzold<sup>1</sup>, T. Andert<sup>2</sup>, M. Hahn<sup>1</sup>, S. W. Asmar<sup>3</sup>, J.–P. Barriot<sup>4</sup>, M. K. Bird<sup>1</sup>, B. Häusler<sup>2</sup>, K. Peter<sup>1</sup>, S. Tellmann<sup>1</sup>, E. Grün<sup>5</sup>, P. R. Weissman<sup>6</sup>, H. Sierks<sup>7</sup>, L. Jorda<sup>8</sup>, R. Gaskell<sup>6</sup>, F. Preusker<sup>9</sup> & F. Scholten<sup>9</sup>

**Cometary nuclei consist mostly of dust and water ice<sup>1</sup>. Previous observations have found nuclei to be low-density and highly porous bodies<sup>2–4</sup>, but have only moderately constrained the range of allowed densities because of the measurement uncertainties. Here we report the precise mass, bulk density, porosity and internal structure of the nucleus of comet 67P/Churyumov–Gerasimenko on the basis of its gravity field. The mass and gravity field are derived from measured spacecraft velocity perturbations at fly-by distances between 10 and 100 kilometres. The gravitational point mass is  $GM = 666.2 \pm 0.2$  cubic metres per second squared, giving a mass  $M = (9,982 \pm 3) \times 10^9$  kilograms. Together with the current estimate of the volume of the nucleus<sup>5</sup>, the average bulk density of the nucleus is  $533 \pm 6$  kilograms per cubic metre. The nucleus appears to be a low-density, highly porous (72–74 per cent) dusty body, similar to that of comet 9P/Tempel 1<sup>2,3</sup>. The most likely composition mix has approximately four times more dust than ice by mass and two times more dust than ice by volume. We conclude that the interior of the nucleus is homogeneous and constant in density on a global scale without large voids. The high porosity seems to be an inherent property of the nucleus material.**

The Rosetta spacecraft arrived at comet 67P/Churyumov–Gerasimenko on 6 August 2014 after a ten year cruise. The final approach manoeuvre was performed at a distance of 100 km from the nucleus. After drifting along with the comet at that distance, the orbit was successfully lowered to 30 km in September 2014 and to 10 km in November 2014.

The goals of the Rosetta Radio Science Investigation experiment (RSI)<sup>6</sup> are to determine the mass of the nucleus, its bulk density and its gravity field in order to constrain the internal structure of the nucleus. At distances of 100 km and 30 km, the nucleus still appears as a point mass expressed as  $GM$  (gravitational constant  $G$  times body mass  $M$ ). It was previously thought that higher moments of the gravity field could only be determined if the spacecraft's orbit about the nucleus were to be lowered to a distance of 10 km or closer<sup>7</sup>. In fact, the higher moments of the gravity field were already sensed and determined at distances below 30 km, a fact which is solely explained by the odd shape of the nucleus.

RSI uses the two-way coherent radio link between the ground station antennas of the ESA's European Space Tracking Network (ESTRACK) and NASA's Deep Space Network (DSN) and the spacecraft. The Doppler shifts of the two radio carrier signals at the X-band (8.4 GHz) and the S-band (2.3 GHz) are examined for changes caused by the perturbing gravitational force of the nucleus acting on the spacecraft. The comet's attracting force changes the trajectory and velocity of the spacecraft, imposing an additional Doppler shift onto the carrier frequencies. The additional Doppler shift can be extracted by predicting and subtracting the unperturbed Doppler shift, assuming that the comet

nucleus is not there. This procedure has been successfully applied to spacecraft fly-bys at planetary bodies in the past<sup>8–10</sup>. Our team developed a software package to process radio tracking fly-by data, which was successfully applied to the fly-bys of Mars Express at Phobos<sup>11,12</sup> and the Rosetta fly-by at asteroid Lutetia<sup>13</sup> (see the Methods).

More than three months' worth of spacecraft radio tracking data between 100 km and 10 km have been processed and solutions of the gravity field generated by the analysis are presented here. Using different initial guesses based on pre-arrival high- and low-mass estimates, consistent solutions of the point mass that are insensitive to the initial guesses are obtained with an average  $GM = (666.2 \pm 0.2) \text{ m}^3 \text{ s}^{-2}$  or  $M = (9,982 \pm 3) \times 10^9 \text{ kg}$  for distances larger than 30 km. The mass of the nucleus is by far the smallest mass ever determined using this technique for a body of the Solar System. The errors of  $GM$  and  $M$  and all other errors given later are one standard deviation ( $1\sigma$ ).

Images from the OSIRIS camera<sup>14</sup> onboard Rosetta were used to determine the bilobate shape and the volume of the nucleus. Two shape models separately derived by the Laboratoire d'Astrophysique de Marseille (LAM) and by DLR Berlin-Adlershof (DLR) give volume estimates of  $(18.8 \pm 0.3) \text{ km}^3$  and  $(18.7 \pm 0.4) \text{ km}^3$ , respectively. These two estimates agree very well within their error bars and are more precise, but are substantially smaller by about 10% than the volume value published earlier<sup>14</sup>. The resulting weighted average volume and bulk density are thus  $(18.7 \pm 0.2) \text{ km}^3$  and  $(533 \pm 6) \text{ kg m}^{-3}$ , respectively. This bulk density value, about half the density of water ice, implies that the nucleus is highly porous.

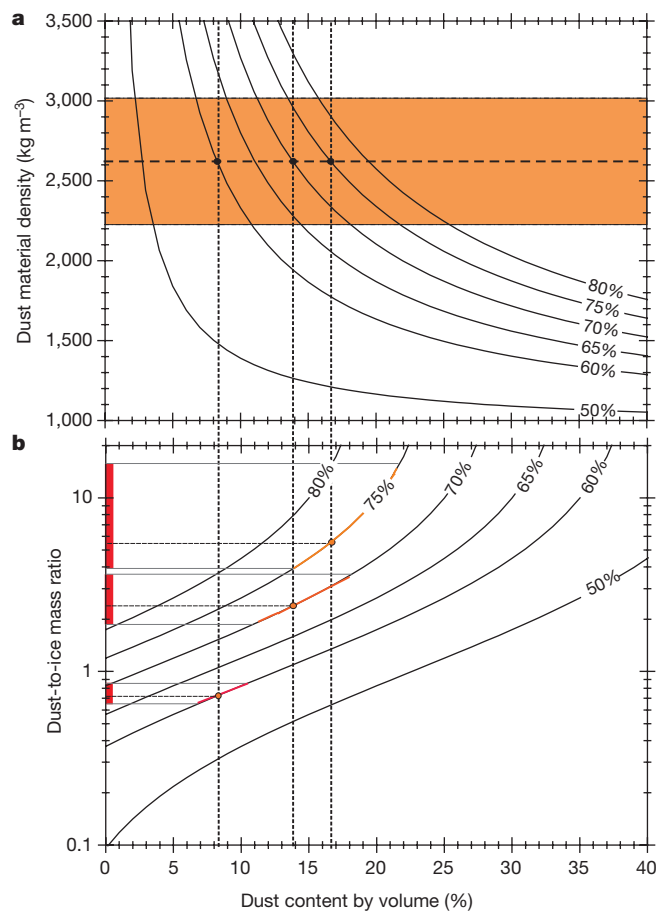
For reasonable dust material densities one can relate the porosity of the nucleus to the dust-to-ice ratio. The cometary dust particles may be extremely porous and fluffy objects, but the silicate material densities are likely to be around  $3,000 \text{ kg m}^{-3}$ , as determined by the Stardust mission<sup>15</sup>. Organic carbonaceous particles may have material densities of  $2,200 \text{ kg m}^{-3}$ . Assuming a ratio of 1:1 for both aggregates would account for an average global material density of  $2,600 \text{ kg m}^{-3}$ , in accordance with earlier estimates<sup>16</sup>.

Figure 1 shows the material density and the dust-to-ice mass ratio as a function of dust content by volume along curves of constant porosity computed from:

$$\rho_{\text{dust}} = \frac{\rho_{\text{bulk}} - \rho_{\text{ice}}(1 - f_{\text{porosity}} - f_{\text{dust}})}{f_{\text{dust}}}$$

where  $\rho_{\text{dust}}$  is the dust material density,  $\rho_{\text{bulk}}$  is the observed average bulk density of  $533 \text{ kg m}^{-3}$ ,  $\rho_{\text{ice}} = 940 \text{ kg m}^{-3}$  is the density of amorphous water ice<sup>17</sup>, and  $f_{\text{porosity}}$  and  $f_{\text{dust}}$  are the fractional contributions of the porosity and the dust to the total volume. The porosity must be greater than 45% for the observed bulk density (a lower limit for the case when  $f_{\text{dust}} > 0$ ).

<sup>1</sup>Rheinisches Institut für Umweltforschung an der Universität zu Köln, Abteilung Planetenforschung, 50931 Köln, Germany. <sup>2</sup>Institut für Raumfahrttechnik und Weltraumnutzung, Universität der Bundeswehr München, 85577 Neubiberg, Germany. <sup>3</sup>Jet Propulsion Laboratory, Caltech, Pasadena, California 91109, USA. <sup>4</sup>Université de la Polynésie Française, Faaa, Tahiti. <sup>5</sup>Max-Planck-Institut für Kernphysik, 69117 Heidelberg, Germany. <sup>6</sup>Planetary Science Institute, 1700 East Fort Lowell Suite 106, Tucson, Arizona 85719, USA. <sup>7</sup>Max-Planck-Institut für Sonnensystemforschung, 37077 Göttingen, Germany. <sup>8</sup>Laboratoire d'Astrophysique de Marseille, 13388 Marseille, France. <sup>9</sup>Institut für Planetenforschung, Deutsches Zentrum für Luft- und Raumfahrt (DLR) Berlin-Adlershof, 12489 Berlin, Germany.



**Figure 1 | The porosity of the nucleus.** Dust material density (a) and dust-to-ice mass ratio (b) as a function of dust content by volume for constant porosity. The horizontal dashed line in a is an average dust material density of a range of reasonable dust material densities of compact dust grains. The curves of constant porosity (given in per cent) give the average dust content by volume for the average material density (vertical dashed lines) and the dust-to-ice mass ratio. The coloured segments of the curves in b are the dust-to-ice mass ratio range for the range of dust material densities in a along each curve of constant porosity.

The 70% and 75% porosity curves in Fig. 1a require a dust volume of 14% and 16.5%, respectively, for the average dust material density of  $2,600 \text{ kg m}^{-3}$ . These porosities yield average dust-to-ice mass ratios of 2.4 and 5.3, respectively, in Fig. 1b. The dust-to-ice mass ratio is 1.9 to 14 for the full range of credible dust material densities between  $2,200 \text{ kg m}^{-3}$  and  $3,000 \text{ kg m}^{-3}$ . This means that a nucleus with 70% to 75% porosity can have up to fourteen times more dust than ice by mass, but can also have up to twice as much ice than dust by volume.

The extremes are the 50% and 80% porosity curves. These porosity cases do not result in reasonable dust-to-ice mass ratios. The dust at the

average dust material density would fill the remaining volume of 20% completely for the 80% porosity case. This would represent a highly porous dusty body with no ice. The 50% porosity case represents a porous icy body with an ice content 18 times higher than for dust by volume. Both cases are considered unrealistic.

The GIADA instrument on Rosetta reports an inferred dust-to-ice mass ratio<sup>18</sup> of  $4 \pm 1$ . Applying this value to the direct surface composition would constrain the surface porosity to 72% to 74%, well within the reasonable range of dust material densities.

If the dust-to-ice mass ratio of 4 is correct<sup>18</sup>, then the nucleus of comet 67P/Churyumov–Gerasimenko appears to be a very porous, very dusty body with a dust content by volume about two times larger than the ice content by volume.

With knowledge of the bulk parameters only, it is still not possible to distinguish clearly between micro- and macro-porosity, that is, between the inherent porosity of fluffy dust particles embedded in the ice matrix or the porosity that would result from large voids within the nucleus body originating from the formation of the nucleus. It has been argued that the micro-porosity of the dust material already accounts for most of the nucleus porosity<sup>16</sup>.

The solution for the gravity field up to degree and order 2, considering all tracking data down to 10 km distance, is listed in Table 1. The tracking data between 10 km and 20 km (from November 2014) were only weakly perturbed by the outgassing with the increasing activity of the nucleus. This resulted in a larger error in the solution for GM compared to the solution for the point mass without appreciable outgassing perturbations (distances  $>30 \text{ km}$ ). Table 1 lists the gravity coefficients derived from the observations and the computed coefficients from both shape models assuming constant density. A similar approach was applied to recent studies<sup>12,19</sup>. All coefficients in Table 1 are fully normalized and referenced to a radius of 2.65 km. The coefficients of degree and order 1 in Table 1 reflect the translation of the individual coordinate systems with respect to a common coordinate system with its origin in the centre of mass defined by the observed gravity field. The translation of the centre of mass of the shape models into the common coordinate system (see Extended Data Table 1) is of the order of the error of the axes of the shape models. Table 1 also directly compares the degree and order-2 coefficients from the RSI observation with those from the two shape models in a common coordinate system. A comparison of the observed and the theoretical gravity coefficients for a homogeneous nucleus in Table 1 reveals that all the  $C_{20}$  and the  $C_{22}$  values agree well within a  $3\sigma$  standard variation.

We therefore conclude that large voids within the nucleus (macro-porosity) can be excluded and that the interior of the nucleus is homogeneous on a global scale, which is in agreement with the CONSERT<sup>4</sup> and OSIRIS observations<sup>20</sup>. The Rosetta CONSERT experiment, which studies the propagation of radio waves in the VHF-band ( $\sim 100 \text{ MHz}$ ) between the Rosetta orbiter and the Philae lander through the nucleus, has concluded that at least the interior of the ‘head’ of the duck-shaped nucleus is homogeneous down to scales of 3 m and that the porosity is 75% to 85% (ref. 19). As stated above, we consider the lower-porosity value to be more likely because porosities larger than

**Table 1 | Comet nucleus gravity field coefficients**

	$l=0$	$l=1$			$l=2$				
	GM ( $\text{m}^3 \text{s}^{-2}$ )	$C_{10}$	$C_{11}$	$S_{11}$	$C_{20}$	$C_{21}$	$C_{22}$	$S_{21}$	$S_{22}$
RSI $>30 \text{ km}^\dagger$	$666.2 \pm 0.2$								
RSI $>10 \text{ km}^\ddagger$	$666.1 \pm 0.4$								
RSI observed		$-0.006 \pm 0.008$	$-0.0001 \pm 0.0024$	$0.0009 \pm 0.006$	$-0.035 \pm 0.002$	$-0.0001 \pm 0.0003$	$0.045 \pm 0.001$	$0.0006 \pm 0.0008$	$-0.0006 \pm 0.0009$
DLR shape		$0.017 \pm 0.001$	$0.007 \pm 0.002$	$-0.0055 \pm 0.0002$	$-0.037 \pm 0.001$	$0.0038 \pm 0.0001$	$0.0450 \pm 0.0004$	$-0.0009 \pm 0.0001$	$0.0006 \pm 0.0001$
LAM shape		$-0.008 \pm 0.001$	$0.006 \pm 0.001$	$-0.0072 \pm 0.0001$	$-0.033 \pm 0.001$	$0.0024 \pm 0.0001$	$0.0444 \pm 0.0004$	$-0.0012 \pm 0.0001$	$-0.0007 \pm 0.0001$

<sup>†</sup>All tracking data between 100 km and 30 km distance; nucleus appears as a point mass.

<sup>‡</sup>All tracking data between 100 km and 10 km.

‘Observed’ means the derived gravity coefficients  $C_m$  and  $S_m$  of degree  $l$  and order  $m$  from all tracking data between 100 km and 10 km; ‘shape’ means computed gravity coefficients from the shape model, assuming constant density. LAM is the shape model from the Laboratoire d’Astrophysique de Marseille; DLR is the shape model from DLR Berlin-Adlershof. All errors are one standard deviation ( $1\sigma$ ).

75% unrealistically constrain the nucleus material to mostly dust with very little ice. The porosity of about 70% to 75% inferred from the mass and the bulk density must therefore be considered an inherent property of the nucleus material.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 27 February; accepted 10 December 2015.**

- Whipple, F. A comet model. I. The acceleration of comet Encke. *Astrophys. J.* **111**, 375–394 (1950).
- A'Hearn, M. F. *et al.* Deep impact: excavating comet Tempel 1. *Science* **310**, 258–264 (2005).
- Ernst, C. M. & Schultz, P. H. Evolution of the Deep Impact flash: implications for the nucleus surface based on laboratory experiments. *Icarus* **191**, 123–133 (2007).
- Kofman, W. *et al.* Properties of the interior of the nucleus of 67P/Churyumov-Gerasimenko revealed by CONSERT radar. *Science* **349**, <http://www.sciencemag.org/content/349/6247/aab0639> (2015).
- Preusker, F. *et al.* Shape model, reference system definition and cartographic mapping standards for comet 67P/Churyumov-Gerasimenko—stereo-photogrammetric analysis of Rosetta/OSIRIS image data. *Astron. Astrophys.* **583**, A33 (2015).
- Pätzold, M. *et al.* Rosetta Radio Science Investigations (RSI). In *Rosetta: ESA's Mission to the Origin of the Solar System* (eds Schulz, R. *et al.*) 537–563 (Springer, 2009).
- Pätzold, M. *et al.* Gravity field determination of a comet nucleus: Rosetta at P/Wirtanen. *Astron. Astrophys.* **375**, 651–660 (2001).
- Anderson, J. D. in *Physical Studies of Minor Planets* (ed. Gehrels, T.) 577–583 (NASA Spec. Publ. SP-267, 1971).
- Anderson, J. D., Lau, E. L., Sjogren, W. L., Schubert, G. & Moore, W. B. Europa's differentiated internal structure: inferences from two Galileo encounters. *Science* **276**, 1236–1239 (1997).
- Yeomans, D. K. *et al.* Estimating the mass of asteroid 253 Mathilde from tracking data during the NEAR flyby. *Science* **278**, 2106–2109 (1997).
- Andert, T. P. *et al.* Precise mass determination and the nature of Phobos. *Geophys. Res. Lett.* **37**, L09202 (2010).
- Pätzold, M. *et al.* Phobos mass determination from the very close flyby of Mars Express in 2010. *Icarus* **229**, 92–98 (2014).
- Pätzold, M. *et al.* Asteroid 21 Lutetia: low mass, high density. *Science* **334**, 491–492 (2011).
- Sierks, H. *et al.* On the nucleus structure and activity of comet 67P/Churyumov-Gerasimenko. *Science* **347**, aaa1044 (2015).
- Brownlee, D. *et al.* Comet 81P/Wild 2 under a microscope. *Science* **314**, 1711–1716 (2006).
- Greenberg, J. M. From comets to meteors. *Earth Moon Planets* **82**, 313–324 (2000).
- Mishima, O., Calvert, L. D. & Lohally, E. An apparently first order transition between two amorphous phases of ice induced by pressure. *Nature* **314**, 76–78 (1985).
- Rotundi, A. *et al.* Dust measurements in the coma of comet 67P/Churyumov-Gerasimenko inbound to the Sun. *Science* **347**, aaa3905 (2015).
- Takahashi, Y. & Scheeres, D. J. Morphology driven density distribution estimation for small bodies. *Icarus* **233**, 179–193 (2014).
- Massironi, M. *et al.* Two independent and primitive envelopes of the bilobate nucleus of comet 67P. *Nature* **526**, 402–405 (2015).

**Acknowledgements** Rosetta is an ESA mission with contributions from its member states and NASA. The Rosetta RSI experiment is funded by the Bundesministerium für Wirtschaft BMWi, Berlin, via the German Space Agency DLR, Bonn, under grants 50QM1401 (RIU-PF) and 50QM1002 (UniBw). J.-P.B. is supported by CNES, Paris. Support for the Multimission Radio Science Support Team is provided by NASA/JPL. We thank everyone involved with the Rosetta mission at ESTEC, ESOC, ESAC and JPL. The RSI team expresses deep appreciation for the critical support provided by the Rosetta SGS at ESAC during the planning and in particular by the ESTRACK and DSN ground station networks during the data acquisition periods. We dedicate this work to the late Claudia Alexander, for her support of RSI over many years.

**Author Contributions** T.A., M.H., J.-P.B., K.P. and S.T. processed the RSI data, S.W.A. was responsible for the data recording at the DSN antennas, H.S., L.J., R.G., F.P. and F.S. provided the nucleus shape models based on OSIRIS images, M.P. is the RSI Principal Investigator and interpreted the data and wrote the paper, B.H. and S.T. are the technical and experiment managers, respectively, M.K.B., B.H., E.G., P.R.W. and all other authors interpreted and discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.P. ([martin.paetzold@uni-koeln.de](mailto:martin.paetzold@uni-koeln.de)).



## METHODS

The Rosetta spacecraft was tracked by ground station antennas of the ESA's European Space Tracking Network (ESTRACK) and NASA's Deep Space Network (DSN). Tracking data from beginning after the orbit insertion on 6 August until mid-November 2014 have been used for the mass and gravity field determination when the spacecraft was at distances to the nucleus between 100 km and 10 km. Times of spacecraft manoeuvring for orbit correction have not been used or modelled. The sampling time at the ground station antenna site was one sample per second.

The received radio carrier frequency at the X-band (8.4 GHz) is compared with a carrier frequency prediction of a spacecraft motion not perturbed by the attracting force of the comet nucleus. This frequency prediction is based on a complex force model taking into account gravitational forces of the Sun, the planets and the large asteroids Ceres, Pallas and Vesta. The planetary and lunar ephemerides DE423 were used. Non-gravitational forces acting on the spacecraft (solar radiation pressure relative to a spacecraft macro model with known optical parameters of each plane and the solar panels and their orientation) and the mass flux on the spacecraft from cometary outgassing were considered. The pressure force by the outgassing was modelled at this stage of the mission by an internal project model. Outgassing was present at all times as measured by the ROSINA experiment<sup>21</sup> but resulted only in weak interactions with the spacecraft and did not affect the mass solution. Also required for a precise frequency prediction is the knowledge of the ground station antenna phase centres and their behaviour under forces like solid Earth tides, plate tectonics and Earth rotation, precession and nutation<sup>22</sup>. Relativistic propagation effects are considered to second order<sup>23</sup>. This frequency prediction is routinely computed for the data processing of the radio science experiments on Mars Express, Venus Express and Rosetta<sup>23,24</sup> and was successfully applied for the Mars Express fly-bys at Phobos and the Rosetta fly-by at asteroid 21 Lutetia<sup>11–13</sup>.

The Doppler shift from the perturbed spacecraft motion caused by the attracting force of the comet nucleus is extracted from the observed carrier frequency shift by subtracting the predicted frequency shift. The difference between the observed perturbed and the predicted Doppler shift is the raw frequency residual.

The raw frequency residuals contain a contribution caused by the propagation of the radio signal through Earth's ionosphere and troposphere. The propagation through the troposphere is mainly affected by the temperature, the atmospheric pressure and partial pressure of the water vapour. These meteorological parameters are routinely recorded at the ground station antenna site.

The tropospheric refraction of the radio ray path in the Earth atmosphere consists of two components: (1) the non-water-vapour component of the atmosphere (the dry component), and (2) the contribution of the highly variable water vapour content of the atmosphere (the wet component). The correction for refraction in Earth's atmosphere is calculated with models for the path delay and mapping functions that project the path delay onto the direction of the signal path for the wet and dry components. The models for the dry component<sup>25</sup>, for the wet component<sup>26</sup>, and the straightforward mapping functions<sup>27</sup> were used to compute the tropospheric correction. The uncertainty in the wet component is much larger than that of the dry component.

The tropospheric correction is subtracted from the raw frequency residual to obtain the tropospherically corrected residual. The high-gain-antenna motion of Rosetta's steerable large antenna dish was corrected by a model successfully applied to the Rosetta–Lutetia fly-by in 2010<sup>13</sup>. The velocity of the spacecraft about the nucleus is of the order of centimetres or tens of centimetres per second and therefore the frequency residuals were filtered with a moving average filter at an integration time of 60 s for noise reduction.

The predicted received frequency was compared to the observed frequency. A weighted least-squares fit to the observed frequency yields a solution for  $GM$  (see Extended Data Fig. 1), the gravity coefficients up to degree and order 2, an adjusted state vector and an adjusted solar radiation pressure constant. The change  $\delta x$  of the initial parameter set  $x$  iteratively aligning the measurement and the model is obtained from:

$$\delta x = (J^T W J + I \alpha)^{-1} J^T W \varepsilon$$

where  $J$  is the Jacobi matrix, containing the partial derivatives of the parameter set  $x$ ,  $W$  is the weighting matrix containing the standard derivatives of the

measurement,  $\varepsilon$  is the difference between model and measurement,  $I$  is the identity matrix and  $\alpha$  is a damping factor. The damping factor serves as a numerical stabilization of the solution against ill-posed parameters<sup>28</sup>. The iterative process is applied until the solution converges, that is, until measurements and models are aligned. The inverse of the term in parentheses is computed using singular value decomposition<sup>29</sup>. The error of each parameter is derived from the diagonal terms of the covariance matrix:

$$Q = (J^T W J)^{-1}$$

Three months' worth of tracking data from 6 August until mid-November was finally processed all together to obtain one solution for all data arcs. The kernels for the SPICE toolkit from NASA's Navigation and Ancillary Information Facility (NAIF) (<https://naif.jpl.nasa.gov/naif/toolkit.html>) used for the computations are listed in Extended Data Table 2.

The least-squares fit yielded also the solutions for the gravity field up to degree and order 2— $C_{10}$ ,  $C_{11}$ ,  $S_{11}$ ,  $C_{20}$ ,  $C_{21}$ ,  $C_{22}$ ,  $S_{21}$  and  $S_{22}$  (Table 1)—an initial state vector (three position and velocity components each) and a solar radiation scale factor for each data arc. In total about 300 parameters were adjusted. Extended Data Table 3 shows the correlations between the fit  $GM$  and the gravity coefficients.

The coordinate system of a preliminary shape model provided by the Rosetta Mission Operations Center defined by a combination of orbit and attitude (CORB and CATT) SPICE kernels listed in Extended Data Table 2 was used. The origin of the coordinate system was located in the centre of mass (COM) of that preliminary shape model and the axes were already aligned along the principal system. The solutions of the degree and order-1 gravity coefficients (Table 1) define the shift of the COM to the true COM (Extended Data Table 1). The reference radius was chosen as  $R_0 = 2.65$  km, slightly larger than the largest axis of the nucleus. A sphere of radius  $R_0$  is fully embracing the nucleus. The COM is shifted from the preliminary COM by:

$$\Delta x = C_{11} R_0 \sqrt{3}$$

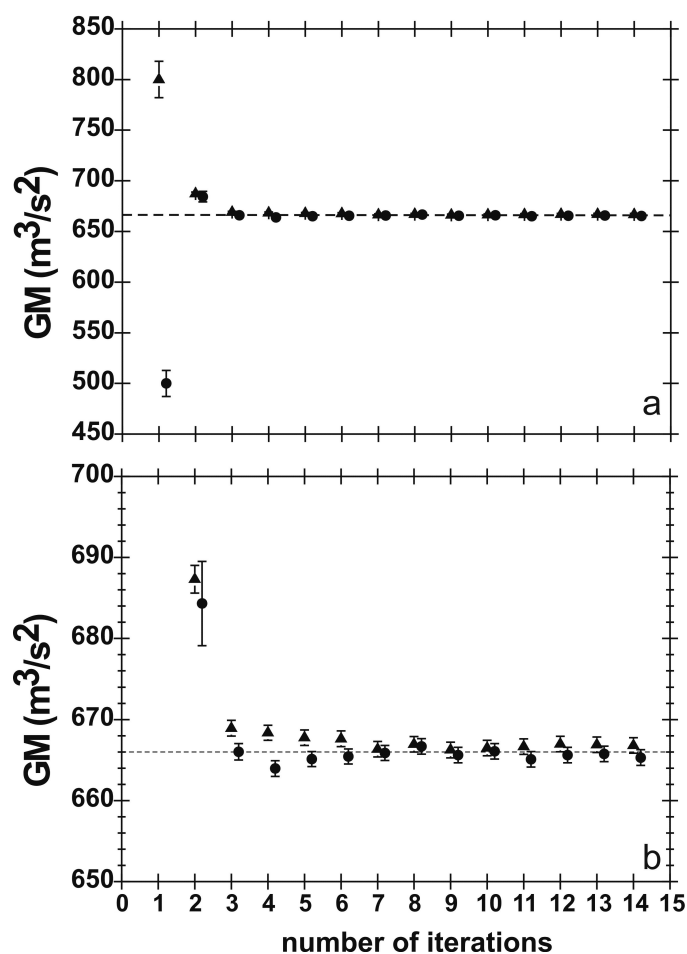
$$\Delta y = S_{11} R_0 \sqrt{3}$$

$$\Delta z = C_{10} R_0 \sqrt{3}$$

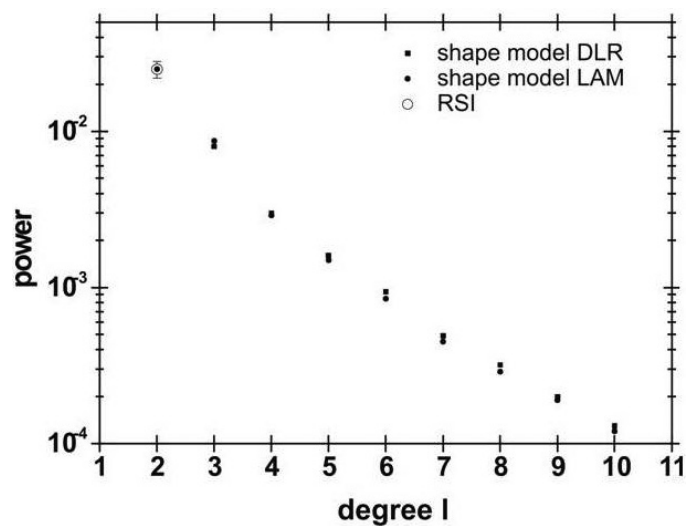
The two shape models from DLR<sup>5</sup> and LAM, derived from OSIRIS images, have their own coordinate system defined by their own COM. These coordinate systems have been transferred into the same coordinate system as the observations. Gravity coefficients up to degree and order 10 were computed by the same method as used for the shape model of Phobos<sup>12</sup> assuming constant density. The computed gravity coefficients for degree and order 1 define different translations to the now common COM (Extended Data Table 1). Extended Data Fig. 2 shows the power spectrum of the observed and computed gravity coefficients.

**Code availability.** We have opted not to make the code available because the code is proprietary and used for other projects.

- Altwegg, K. *et al.* 67P/Churyumov-Gerasimenko, a Jupiter family comet with a high D/H ratio. *Science* **347**, 1261592 (2015).
- McCarthy, D. D. & Petit, G. *IERS Technical Note* 32. <http://www.iers.org/IERS/EN/Publications/TechnicalNotes/tn32.html?nn=94912> (Verlag des Bundesamts für Kartographie und Geodäsie, IERS Conventions, 2004).
- Häusler, B. *et al.* Radio science investigations by VeRa onboard the Venus Express spacecraft. *Planet. Space Sci.* **54**, 1315–1335 (2006).
- Pätzold, M. *et al.* in *Mars Express: The Scientific Investigations* ESA SP-1291, 217–248 (European Space Agency, 2009).
- Saastamoinen, J. in *The Use of Artificial Satellites for Geodesy* (eds Henriksen, S. W. *et al.*) 247–252 (American Geophysical Union, 1972).
- Ifadis, I. I. *The atmospheric delay of radio waves: modeling the elevation dependence on a global scale*. Technical Report No. 38L (School of Electrical and Computer Engineering, Chalmers University of Technology, 1986).
- Chao, C. C. *A model for tropospheric calibration from daily surface and radiosonde balloon measurements*. Technical Memorandum 391–350 (Jet Propulsion Laboratory, 1972).
- Aster, R. C., Borchers, B. & Thurber, C. *Parameter Estimation and Inverse Problems* (Elsevier Academic Press, 2005).
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in Fortran* (Cambridge Univ. Press, 1986).



**Extended Data Figure 1 | Gravitational mass solutions.** **a**,  $GM$  solutions after each iteration step for two different starting values estimated from pre-arrival considerations. For the circle symbols, the initial value is  $800 \text{ m}^3 \text{ s}^{-2}$ ; for the triangle symbols it is  $500 \text{ m}^3 \text{ s}^{-2}$ . The final solutions converge to an average  $GM$  of  $666.2 \pm 0.2 \text{ m}^3 \text{ s}^{-2}$  for both runs (tracking data for distances larger than 30 km). **b**, As for **a**, but with a limited ordinate and not showing the values of the initial values. The error bars are one standard deviation ( $1\sigma$ ).



**Extended Data Figure 2 | Gravity coefficients power spectrum.** The power spectrum of the computed gravity coefficients from the shape models for degree  $l = 2$  to 10. The degree-2 power observed by RSI is the open data point with error bar ( $1\sigma$ ), which agrees very well with degree 2 from the shape models.



Extended Data Table 1 | Translation of the preliminary coordinate systems to the common coordinate system

	RSI	Shape models	
		DLR	LAM
$\Delta x$ (m)	0 +/- 11	32 +/- 9	28 +/- 5
$\Delta y$ (m)	4 +/- 28	-25 +/- 1	-33 +/- 1
$\Delta z$ (m)	-28 +/- 37	78 +/- 6	-37 +/- 5

Extended Data Table 2 | SPICE kernels used for gravity field computation

Kernel Type	Kernel name
C-G ephemeris kernel (SPK)	CORB_DV_074_01_____00128.BSP
Rosetta orbit/ephemeris kernel (SPK)	RORB_DV_074_01_____00128.BSP
Rosetta frame kernel (FK)	ROS_V24.TF
C-G Attitude kernel (CK)	CATT_DV_074_01_____00128.BC
Rosetta Spacecraft Clock Kernel (TSC)	ROS_141113_STEP.TSC

Extended Data Table 3 | Correlation matrix of the fit parameters *GM* and the gravity coefficients up to order and degree 2

	GM	C10	C11	S11	C20	C21	C22	S21	S22
GM	1.00	0.15	0.00	-0.14	0.05	0.07	0.01	-0.09	0.02
C10	0.15	1.00	-0.05	-0.06	0.08	-0.02	0.01	-0.03	0.00
C11	0.00	-0.05	1.00	-0.03	0.07	0.05	0.04	0.29	0.01
S11	-0.14	-0.06	-0.03	1.00	-0.04	-0.13	0.03	0.10	-0.02
C20	0.05	0.08	0.07	-0.04	1.00	-0.06	-0.13	-0.11	0.05
C21	0.07	-0.02	0.05	-0.13	-0.06	1.00	0.14	0.03	-0.01
C22	0.01	0.01	0.04	0.03	-0.13	0.14	1.00	0.10	-0.01
S21	-0.09	-0.03	0.29	0.10	-0.11	0.03	0.10	1.00	0.01
S22	0.02	0.00	0.01	-0.02	0.05	-0.01	-0.01	0.01	1.00



# Optical attosecond pulses and tracking the nonlinear response of bound electrons

M. Th. Hassan<sup>1\*</sup>, T. T. Luu<sup>1\*</sup>, A. Moulet<sup>1</sup>, O. Raskazovskaya<sup>2</sup>, P. Zhokhov<sup>3,4</sup>, M. Garg<sup>1</sup>, N. Karpowicz<sup>1</sup>, A. M. Zheltikov<sup>3,4</sup>, V. Pervak<sup>2</sup>, F. Krausz<sup>1,2</sup> & E. Goulielmakis<sup>1</sup>

**The time it takes a bound electron to respond to the electromagnetic force of light sets a fundamental speed limit on the dynamic control of matter and electromagnetic signal processing. Time-integrated measurements of the nonlinear refractive index<sup>1</sup> of matter indicate that the nonlinear response of bound electrons to optical fields is not instantaneous; however, a complete spectral characterization of the nonlinear susceptibility tensors<sup>2</sup>—which is essential to deduce the temporal response of a medium to arbitrary driving forces using spectral measurements—has not yet been achieved. With the establishment of attosecond chronoscopy<sup>3–5</sup>, the impulsive response of positive-energy electrons to electromagnetic fields has been explored through ionization of atoms<sup>6</sup> and solids<sup>7</sup> by an extreme-ultraviolet attosecond pulse<sup>8</sup> or by strong near-infrared fields<sup>9–11</sup>. However, none of the attosecond studies carried out so far have provided direct access to the nonlinear response of bound electrons. Here we demonstrate that intense optical attosecond pulses synthesized in the visible and nearby spectral ranges allow sub-femtosecond control and metrology of bound-electron dynamics. Vacuum ultraviolet spectra emanating from krypton atoms, exposed to intense waveform-controlled optical attosecond pulses, reveal a finite nonlinear response time of bound electrons of up to 115 attoseconds, which is sensitive to and controllable by the super-octave optical field. Our study could enable new spectroscopies of bound electrons in atomic, molecular or lattice potentials of solids<sup>12</sup>, as well as light-based electronics operating on sub-femtosecond timescales and at petahertz rates<sup>13–15</sup>.**

Intramolecular atomic motion dynamically distorts the interatomic electron cloud and thereby gives rise to a transient electronic polarizability<sup>16</sup> and a finite nonlinear response of matter to light fields. Time-resolved spectroscopies can now probe this response with femtosecond lasers<sup>17</sup>. A finite nonlinear response of matter to external fields can also emerge in the absence of nuclear degrees of freedom, purely as a result of electron dynamics caused by coherent excitation of electronic states and concomitant spatiotemporal evolution of the electron cloud<sup>18,19</sup>. However, quantum mechanics implies that these dynamics occur on an attosecond (as) timescale of  $\Delta t \approx \hbar/\Delta E \approx 400\text{--}800$  as (in which  $\Delta E$  is the energy spacing between the relevant electronic levels, typically about 5–10 eV, and  $\hbar$  is Planck's constant). Consequently, it has long been conjectured that the build-up of the nonlinear polarization of bound electrons under non-resonant driving by light fields is instantaneous<sup>20</sup>. For electron dynamics to influence substantially the nonlinear response of a medium to an optical field and to enable its real-time tracing, both induction of electronic coherence and nonlinear probing of concomitant polarization dynamics must occur within time intervals shorter or comparable to  $\Delta t$ . In the visible and nearby spectral ranges, where bound electrons are typically probed and the required field strengths for nonlinear studies can be most readily produced, this condition implies a light pulse with its energy substantially being

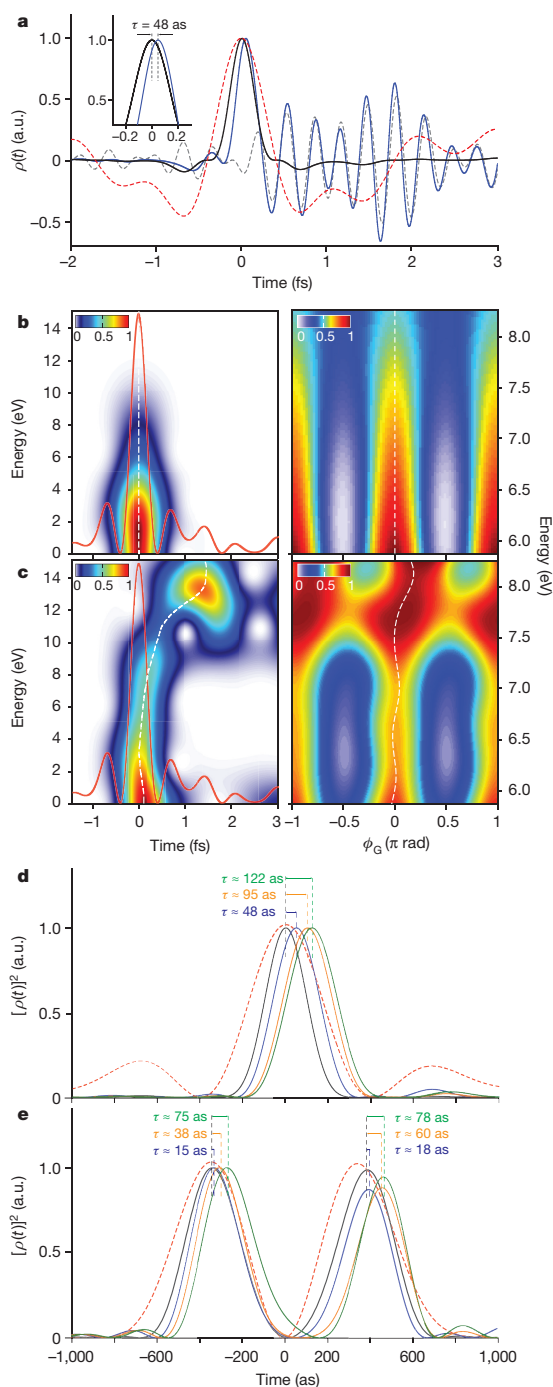
confined to a half wave cycle. We refer to such a pulse as an optical attosecond pulse.

To illustrate how such a pulse can enable the study of the dynamics of the bound-electron response, we theoretically investigate the interaction of an intense ( $4 \times 10^{13} \text{ W cm}^{-2}$ ) half-cycle optical field (Fig. 1a, red dashed line), identical to that used in our experiments described below, with a system lacking nuclear degrees of freedom—a krypton (Kr) atom. We calculate the induced nonlinear dipole moment  $\rho(t)$  by numerical integration of the three-dimensional time-dependent Schrödinger equation (TDSE; Fig. 1a, blue line;  $\rho_{\text{TDSE}}(t)$ ) and by using an adiabatic model describing Kr (Fig. 1a, black line;  $\rho_{\text{ad}}(t)$ ; see Methods). The instantaneous response that is inherent to the adiabatic model is manifested by the accurate synchronization between the optical field  $E(t)$  (Fig. 1a, red dashed line) and the nonlinear dipole moment  $\rho_{\text{ad}}(t)$  (Fig. 1a, black line) induced in the system, and is further corroborated by the synchronous emission of the  $\rho_{\text{ad}}(\omega)$  spectrum (where  $\omega$  is frequency) obtained from time–frequency analysis of the dipole (Fig. 1b, left panel). In contrast, when we account for the detailed electronic structure of Kr when using the TDSE to simulate the interaction with the ultrafast optical field, coherent oscillations of the electron cloud (Fig. 1a, grey dashed line) markedly affect the nonlinear dipole  $\rho_{\text{TDSE}}(t)$ . The leading-edge and maximum of  $\rho_{\text{TDSE}}(t)$  are noticeably retarded, relative to those of the instantaneous response  $\rho_{\text{ad}}(t)$  and the driving electric field  $E(t)$ . The time–frequency analysis of  $\rho_{\text{TDSE}}(t)$  (Fig. 1c, left panel) reveals an attosecond-scale, non-uniform retardation (group delay dispersion) of the nonlinear emission (Fig. 1c, left panel, white dashed line) that is evident across the entire dipole spectrum and extends to the resonant area (10–15 eV) of Kr. Here, the nonlinear response is dominated by long-lasting coherences that typically evolve over substantially longer timescales (Fig. 1a, grey dashed line). In contrast, the nonlinear emission at lower energies ( $<9$  eV) is virtually confined within the time interval of the interaction of the optical field with Kr. Hence, this low-energy spectral range offers an adequate basis for exploring the dynamics of the nonlinear response of the electron cloud to the optical driver.

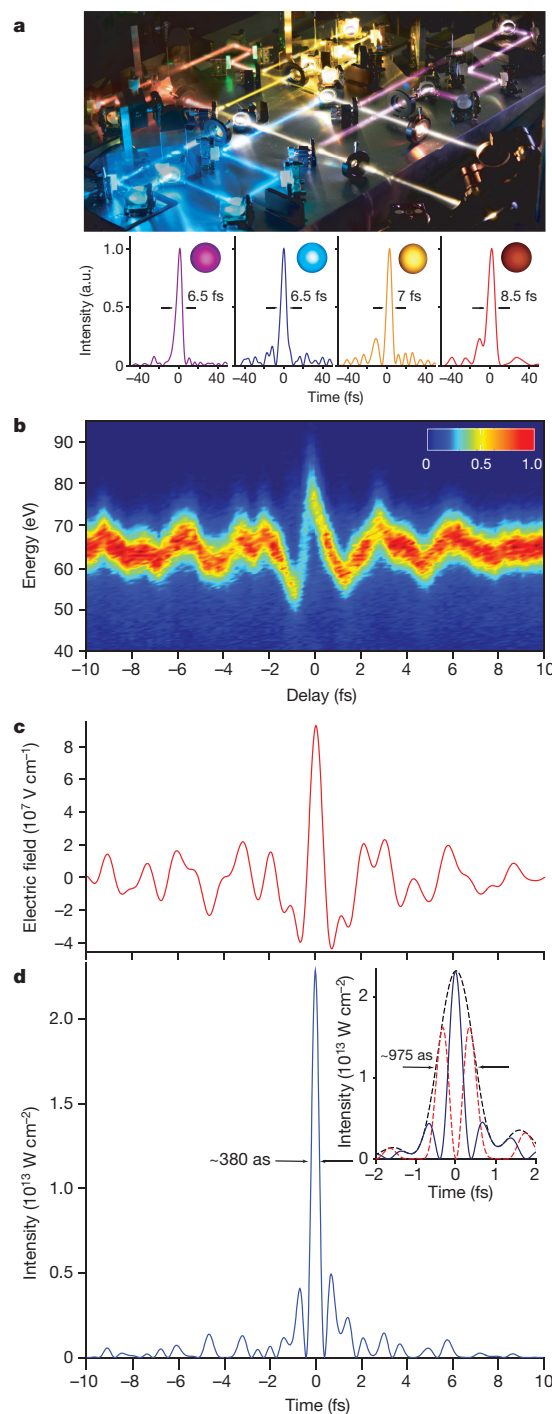
A comparison between the Fourier-filtered ( $<9$  eV) nonlinear dipoles  $\rho_{\text{ad}}(t)$  and  $\rho_{\text{TDSE}}(t)$  predicted by the adiabatic model (Fig. 1d, black line) and the TDSE model (Fig. 1d, blue line), respectively, reveals a finite relative delay of  $\tau \approx 48$  as (Fig. 1a inset). This delay represents a substantial fraction of the time span of the induced dipole  $\rho_{\text{TDSE}}(t)$ , which is about 250 as. The observed retardation is sensitive to, and is monotonically increasing with, applied field strength (compare the blue, orange and green lines in Fig. 1d); with an optical-field intensity of  $8 \times 10^{13} \text{ W cm}^{-2}$ , it reaches a value of  $\tau \approx 122$  as. Figure 1e shows a similar sensitivity of the dynamics of the dipole when the field intensity is adjusted by varying the global phase  $\phi_G$  of the driving field (justification for why the term carrier-envelope phase (CEP) is inappropriate for multi-octave waveforms is given in Methods).

<sup>1</sup>Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, D-85748 Garching, Germany. <sup>2</sup>Department für Physik, Ludwig-Maximilians-Universität, Am Coulombwall 1, D-85748 Garching, Germany. <sup>3</sup>Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843, USA. <sup>4</sup>Physics Department, International Laser Center, M.V. Lomonosov Moscow State University, 119992 Moscow, Russia.

\*These authors contributed equally to this work.



**Figure 1 | Nonlinear response of bound electrons of Kr to an optical attosecond pulse.** **a**, Nonlinear dipole moment  $\rho(t)$  of Kr atoms subjected to an intense half-cycle optical field (dashed red line) calculated using the adiabatic ( $\rho_{ad}(t)$ , black line) and the TDSE ( $\rho_{TDSE}(t)$ , blue line) model. The dashed grey line represents the bound-electron response in the energy range (10–15 eV) that is calculated using the TDSE model. The inset shows a close-up of the peak at time  $t = 0$ , highlighting the delay  $\tau$  between the predictions of the two models. **b, c**, Time–frequency analysis of the nonlinear dipoles  $\rho_{ad}(t)$  (**b**) and  $\rho_{TDSE}(t)$  (**c**) (left panels); simulated nonlinear dipole spectra (5.7–8.2 eV) as a function of the global phase  $\phi_G$  of the optical attosecond pulse predicted by the adiabatic (**b**) and TDSE (**c**) Kr models (right panels). The red line represents the instantaneous intensity of the driving field. The dashed white lines represent the centre of mass along the  $x$  axis of each plot. The colour scale represents spectral intensity in arbitrary units (a.u.). **d, e**, Low-pass-filtered (0–8 eV) normalized nonlinear dipoles for three peak intensities of the driving field:  $4 \times 10^{13} \text{ W cm}^{-2}$  (blue lines),  $6 \times 10^{13} \text{ W cm}^{-2}$  (orange lines) and  $8 \times 10^{13} \text{ W cm}^{-2}$  (green lines) for global phases  $\phi_G \approx 0$  (**d**) and  $\phi_G \approx \pi/2$  rad (**e**).



**Figure 2 | Synthesis of an optical attosecond pulse.** **a**, Photograph of an attosecond light-field synthesizer (top). The beams (artificially visualized) of ultra-wide-band pulses (about 1.1–4.6 eV) are divided by dichroic beam-splitters into four, almost equally wide, spectral bands. The pulses in each band are compressed so that they have durations of a few femtoseconds (the temporal intensity profiles are shown in the bottom panels), before they are spatiotemporally superimposed to yield a single beam/pulse at the exit of the apparatus. The insets in the bottom panels show representative beam profiles in the far field. **b**, Attosecond-streaking spectrogram of an optical attosecond pulse. The colour bar represents the yield of photoelectron counts in arbitrary units. **c, d**, Evaluated electric field (**c**) and instantaneous intensity profile (**d**; blue line) of the pulse. The intensity profile has a FWHM duration of approximately 380 as. The inset of **d** shows a close-up of this instantaneous intensity profile (blue line,  $\phi_G \approx 0$ ) along with that of the same pulse, but with  $\phi_G \approx \pi/2$  rad (dashed red line), and their common intensity envelope (dashed black line), which has a FWHM duration of approximately 975 as.

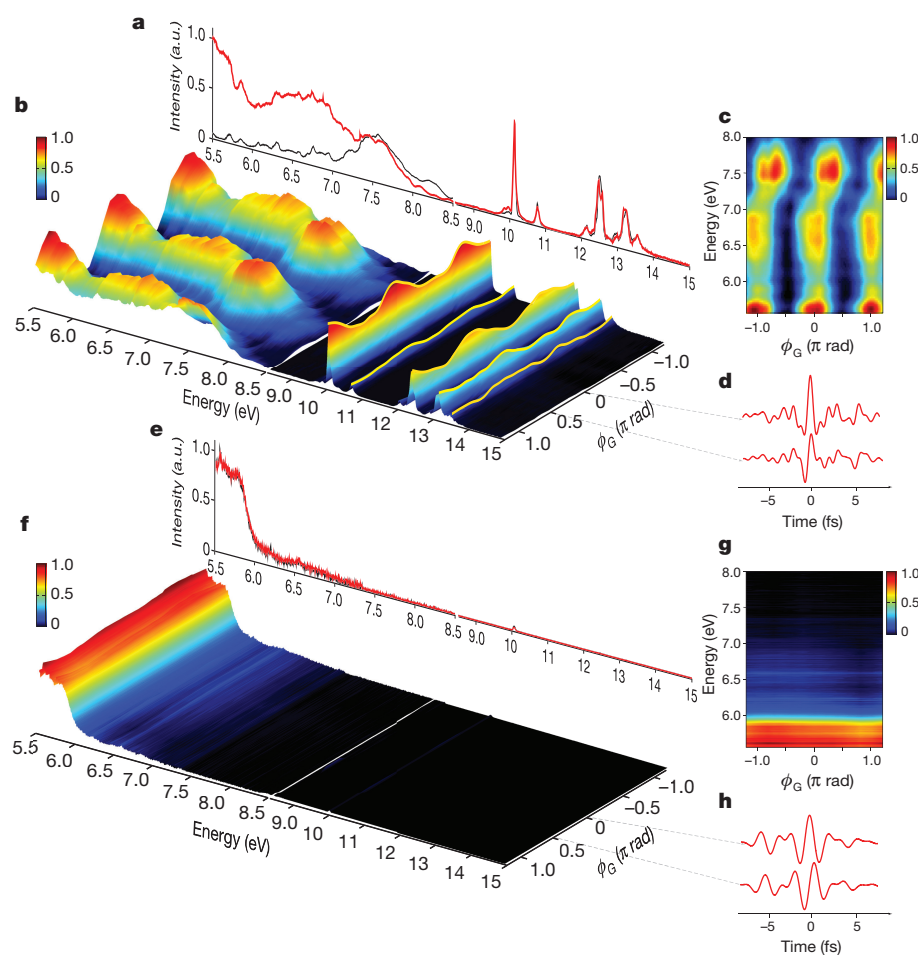
Phase-gating<sup>21</sup> methods have successfully been used for tracing ionized-electron dynamics. To determine whether these methods can be extended to trace bound-electron dynamics with optical attosecond pulses, we simulate nonlinear dipole spectra as a function of the global phase  $\phi_G$  of the optical attosecond pulse using the adiabatic model ( $\rho_{\text{ad}}(\omega, \phi_G)$ , Fig. 1b, right panel) and the TDSE model ( $\rho_{\text{TDSE}}(\omega, \phi_G)$ , Fig. 1c, right panel). The instantaneity of the nonlinear response underlying the adiabatic model (Fig. 1b, left panel) results in uniform spectral modulations of the emitted radiation with  $\phi_G$  (Fig. 1b, right panel, shown for the range 5.7–8.2 eV to ease comparison with experiments). Indeed, the noticeable deviation between the instantaneous and TDSE time-domain nonlinear dipoles (Fig. 1d, e), and the dependence of this deviation on  $\phi_G$ , give rise to easily discernible features in the corresponding spectrogram (Fig. 1c, left panel). These features are the manifestation of the delayed nonlinear response and include the non-synchronous amplitude modulation of the emitted frequencies with global phase (highlighted by the white dashed line) and the non-monotonic variation of the amplitude of the spectral emission along the energy axis.

We describe the nonlinear response of bound electrons in Kr with a simple heuristic model that expresses the nonlinear dipole  $\rho(t, \phi_G)$  as a sum of instantaneous and delayed nonlinearities:

$$\rho(t, \phi_G) = aE^3(t, \phi_G) + bE^5(t, \phi_G) + cE^5(t - dt, \phi_G) \quad (1)$$

where  $a$ ,  $b$  and  $c$  are coefficients,  $dt$  represents a delay and  $E(t, \phi_G)$  is the electric field of the optical attosecond pulse for a global phase  $\phi_G$  (see Methods). Provided  $E(t, \phi_G)$  is known, equation (1) allows retrieval of the complete nonlinear dipole  $\rho(t, \phi_G)$  from the numerical reconstruction of the corresponding  $\phi_G$  spectrogram (Fig. 1c, right panel). This procedure underlines the suitability of the model for retrieving the nonlinear bound-electron response, and is instrumental for unveiling such dynamics from the experiments presented below.

To experimentally study the sub-femtosecond nonlinear response of bound electrons to light fields, we synthesized optical attosecond pulses. Two major advances of our synthesis method with respect to previously used methods<sup>22,23</sup> have enabled the confinement and manipulation of optical pulses to  $<1$  fs. First, the development of a second-generation light-field synthesizer (Fig. 2a), which extends optical synthesis to the deep ultraviolet range and manipulates pulses with spectra extending over more than two optical octaves over the visible and neighbouring ranges (about 1.1–4.6 eV), was crucial. Second, and equally important, was the realization of a broadband and virtually dispersion-free spectral equalization of the synthesized pulses, to give pulses with twice the initial bandwidth (see Methods section ‘Optical attosecond pulse synthesis’). These advances allow synthesis of optical attosecond pulses, which are highly energetic (about 50  $\mu\text{J}$ ) and several



**Figure 3 | Attosecond control of bound electrons in Kr.** **a, e**, VUV spectra generated in neutral Kr atoms driven by an optical attosecond pulse (**a**) and a single-cycle pulse (**e**) of the same peak intensity (about  $5 \times 10^{13} \text{ W cm}^{-2}$ ), at global phase settings of  $\phi_G \approx 0$  (red line) and  $\phi_G \approx \pi/2$  rad (black line). **b, f**, Spectrograms composed of 25 VUV spectra

(about 5.5–15 eV) recorded as a function of  $\phi_G$  for an optical attosecond pulse (**b**) and a single-cycle pulse (**f**). **c, g**, Sections (about 5.5–8 eV) of the global-phase spectrograms in **b** and **f**, respectively. **d, h**, Representative driving waveforms at  $\phi_G \approx 0$  (**d**) and  $\phi_G \approx \pi/2$  rad (**h**). The colour scales in **b, c, f** and **g** indicate intensity (in arbitrary units).



orders of magnitude more intense (about 50 GW) than those obtainable with current extreme-ultraviolet (EUV) technology<sup>24</sup>.

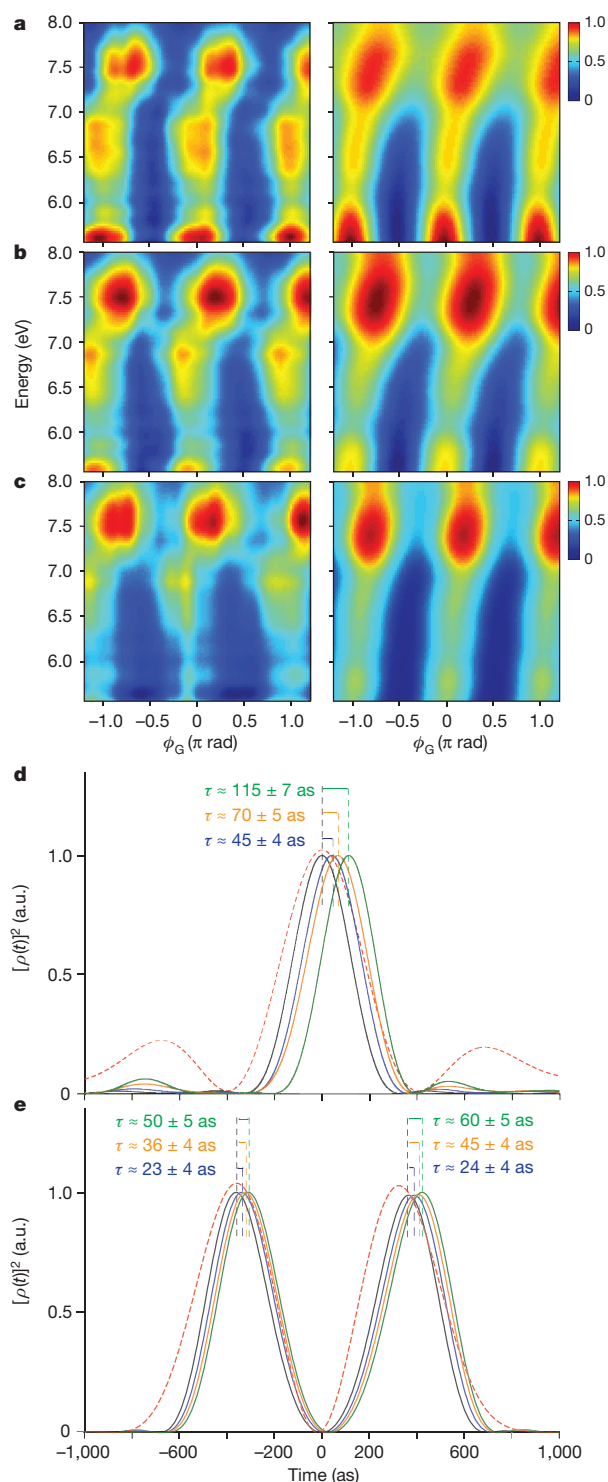
Figure 2b shows a representative streaking spectrogram of an optical attosecond pulse. The field waveform (Fig. 2c) retrieved from the data reveals a nearly isolated, half-cycle peak (centroid wavelength of approximately 530 nm), which has an intensity (Fig. 2d) approximately five times that of the adjacent peaks and carries more than about 50% of the energy of the waveform. The instantaneous intensity profile of the dominant half-cycle in Fig. 2d has a full-width at half-maximum (FWHM) duration of approximately 380 as, whereas an evaluation based on the conventional definition of the intensity envelope (Fig. 2d inset, dashed black line) yields a FWHM of approximately 975 as.

In our experiments (see Extended Data Fig. 1b), we probe the nonlinear polarization response of Kr atoms to optical attosecond pulses by recording the generated vacuum-ultraviolet (VUV) spectra. As shown in Fig. 3a, the spectra are quasi-continuous in the 5.5–10 eV range and exhibit discrete peaks coinciding with Kr excited-state energies in the 10–15 eV range. Variation of the global phase  $\phi_G$  of the pulses yields a pronounced broadband modulation of the VUV emission (Fig. 3b), verifying our TDSE model predictions (Fig. 1c, right panel). In contrast, a pulse synthesized to consist of approximately a single optical cycle of commensurable intensity fails to generate broadband nonlinear excitation or its  $\phi_G$  control (Fig. 3e–g). These results clearly demonstrate that in the absence of the extreme nonlinearities that arise under conditions of ionization, optical attosecond pulses are the key to manipulating bound electrons on a sub-femtosecond timescale.

The non-uniform modulation of the amplitude of the emitted dipole spectra (Fig. 3c) with  $\phi_G$  and the uneven distribution of spectral amplitudes along the energy axis differ from the predictions of the adiabatic model for the same driving waveform (Fig. 1b, right panel), but resemble key features of the spectrograms constructed using TDSE simulations (Fig. 1c, right panel). These features offer experimental evidence of the non-instantaneous character of the response of the bound electrons of Kr in our experiments.

To apply the methodology presented above to retrieve the dynamics of the bound-electronic response, we supplemented our measurements with attosecond streaking of the driving field  $E(\tau, \phi_G)$ . By studying the dependence of the VUV emission as a function of the intensity of the optical driver, we verified that bound-electronic nonlinearities dominate the response (see Methods). Figure 4 shows the measured (left panels) and the corresponding reconstructed (right panels) spectrograms recorded at gradually increasing optical-field intensities: about  $5 \times 10^{13} \text{ W cm}^{-2}$  (Fig. 4a); about  $7 \times 10^{13} \text{ W cm}^{-2}$  (Fig. 4b); and about  $9 \times 10^{13} \text{ W cm}^{-2}$  (Fig. 4c). The retrieved nonlinear dipoles (Fig. 4d) exhibit delays with respect to the instantaneous dipole (black line) and driving field  $E(\tau, 0)$  (red dashed line) of  $\tau \approx 45 \pm 4$  as (Fig. 4a),  $\tau \approx 70 \pm 5$  as (Fig. 4b) and  $\tau \approx 115 \pm 7$  as (Fig. 4c). The data in Fig. 4d verify the attosecond control of the bound-electronic response induced by the field, and emphasize the non-trivial character of the induced dynamics, which is manifested by a non-uniform delay (chirp). The latter is highlighted by the delay values for consecutive half-field-cycles shown in Fig. 4e. These findings, which are in excellent agreement with the predictions of the TDSE simulations performed for identical excitation fields, provide conclusive evidence for the feasibility of tracing and control of the nonlinear response of bound electrons on a sub-femtosecond timescale with high fidelity.

The findings of this study are compatible with an intuitive physical picture of the delayed nonlinear response, which we can now extend from the femtosecond molecular scale<sup>16</sup> to attosecond electron dynamics. The optical-field-induced superposition of the ground state and excited states—which are typically more polarizable because the electron resides further from the atomic core—gives rise to an increase in the total nonlinear polarizability of the electron cloud during its interaction with the field. But this superposition requires a finite time to evolve, resulting in a delay in the nonlinear response of the system.



**Figure 4 | Sub-femtosecond, delayed nonlinear response of bound electrons in Kr.** **a–c**, Global-phase spectrograms of the optical attosecond pulse recorded at peak-intensity settings of about  $5 \times 10^{13} \text{ W cm}^{-2}$  (**a**),  $7 \times 10^{13} \text{ W cm}^{-2}$  (**b**) and  $9 \times 10^{13} \text{ W cm}^{-2}$  (**c**) are shown in the left panels. Corresponding reconstructed spectrograms based on equation (1) are shown in the right panels. The colour bar represents spectral intensity in arbitrary units. **d, e**, Low-pass-filtered (0–8 eV) nonlinear dipoles obtained using the intensity settings in **a–c** (dark blue, orange, green lines, respectively) are shown along with the instantaneous response (black line) simulated for global-phase settings of the optical attosecond pulse of  $\phi_G \approx 0$  (**d**) and  $\phi_G \approx \pi/2$  rad (**e**). Standard errors of the mean for the delays ( $\tau$ ) indicated are evaluated from the reconstruction of three data sets recorded under identical experimental conditions. The dashed red lines are the normalized instantaneous intensities of the input electric fields.

Because both the amplitude of this superposition and the Stark shift<sup>22</sup> of the induced electronic dipoles scale with field strength, the delay in the nonlinear response will also increase, in agreement with our theoretical and experimental studies.

It should be possible to extend the approach demonstrated here to the condensed phase. This will allow the electronic response in the bulk of complex systems subjected to strong fields to the probed<sup>14,25,26</sup>, and will enable the possibility to trace electronic dephasing in such systems. By providing a probe, which is as short as the lifetime and dephasing times of core-hole electronic states in solids generated under attosecond X-ray excitation<sup>15</sup>, our tools could provide new ways to explore the nonlinear properties of X-ray excited solids. These novel spectroscopic capabilities may turn out to be instrumental—among others—to exploring the ultimate frontiers of electromagnetic signal processing in optical and X-ray regimes.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 10 June; accepted 3 December 2015.**

1. Sheik-Bahae, M., Hutchings, D. C., Hagan, D. J. & Van Stryland, E. W. Dispersion of bound electron nonlinear refraction in solids. *IEEE J. Quantum Electron.* **27**, 1296–1309 (1991).
2. Boyd, R. W. *Nonlinear Optics* Ch. 1 (Academic Press, 1992).
3. Hentschel, M. *et al.* Attosecond metrology. *Nature* **414**, 509–513 (2001).
4. Niikura, H. *et al.* Sub-laser-cycle electron pulses for probing molecular dynamics. *Nature* **417**, 917–922 (2002).
5. Corkum, P. B. & Krausz, F. Attosecond science. *Nature Phys.* **3**, 381–387 (2007).
6. Schultze, M. *et al.* Delay in photoemission. *Science* **328**, 1658–1662 (2010).
7. Cavalieri, A. L. *et al.* Attosecond spectroscopy in condensed matter. *Nature* **449**, 1029–1032 (2007).
8. Goulielmakis, E. *et al.* Single-cycle nonlinear optics. *Science* **320**, 1614–1617 (2008).
9. Uiberacker, M. *et al.* Attosecond real-time observation of electron tunnelling in atoms. *Nature* **446**, 627–632 (2007).
10. Eckle, P. *et al.* Attosecond ionization and tunneling delay time measurements in helium. *Science* **322**, 1525–1529 (2008).
11. Shafir, D. *et al.* Resolving the time when an electron exits a tunnelling barrier. *Nature* **485**, 343–346 (2012).
12. Ivanov, M. & Smirnova, O. Opportunities for sub-laser-cycle spectroscopy in condensed phase. *Chem. Phys.* **414**, 3–9 (2013).
13. Krüger, M., Schenk, M. & Hommelhoff, P. Attosecond control of electrons emitted from a nanoscale metal tip. *Nature* **475**, 78–81 (2011).
14. Schiffrin, A. *et al.* Optical-field-induced current in dielectrics. *Nature* **493**, 70–74 (2013).
15. Schultze, M. *et al.* Controlling dielectrics with the electric field of light. *Nature* **493**, 75–78 (2013).
16. Weiner, A. M. *Ultrafast Optics* Ch. 6 (John Wiley & Sons, 2009).
17. Zewail, A. H. Femtochemistry: atomic-scale dynamics of the chemical bond. *J. Phys. Chem. A* **104**, 5660–5694 (2000).
18. Goulielmakis, E. *et al.* Real-time observation of valence electron motion. *Nature* **466**, 739–743 (2010).
19. Smirnova, O. *et al.* High harmonic interferometry of multi-electron dynamics in molecules. *Nature* **460**, 972–977 (2009).
20. Zewail, A. H. *Femtochemistry: Ultrafast Dynamics of the Chemical Bond* Vol. 1 (World Scientific, 1994).
21. Dudovich, N. *et al.* Measuring and controlling the birth of attosecond XUV pulses. *Nature Phys.* **2**, 781–786 (2006).
22. Wirth, A. *et al.* Synthesized light transients. *Science* **334**, 195–200 (2011).
23. Hassan, M. Th. *et al.* Invited Article: attosecond photonics: synthesis and control of light transients. *Rev. Sci. Instrum.* **83**, 111301 (2012).
24. Sansone, G., Poletto, L. & Nisoli, M. High-energy attosecond light sources. *Nature Photon.* **5**, 655–663 (2011).
25. Vampa, G. *et al.* Linking high harmonics from gases and solids. *Nature* **522**, 462–464 (2015).
26. Hohenleutner, M. *et al.* Real-time observation of interfering crystal electrons in high-harmonic generation. *Nature* **523**, 572–575 (2015).

**Acknowledgements** This work was supported by European Research Council grant (Attoelectronics-258501), the Deutsche Forschungsgemeinschaft Cluster of Excellence: Munich Centre for Advanced Photonics (<http://www.munich-photonics.de>), the Max Planck Society and the European Research Training Networks ATTOFEL and MEDEA, the Russian Foundation for Basic Research (projects numbers 13-02-01465, 13-02-92115 and 13-04-40335) and the Welch Foundation (grant number A-1801).

**Author Contributions** M.Th.H., T.T.L., A.M. and M.G. conducted the experiments; O.R. and V.P. developed the dispersive optics for the experiments; E.G. planned the experiments and supervised the project; M.Th.H., T.T.L., P.Z., A.M. and N.K. conducted the simulations; and M.Th.H., T.T.L., A.M., A.M.Z., F.K. and E.G. interpreted the data and contributed to the preparation of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G. ([elgo@mpq.mpg.de](mailto:elgo@mpq.mpg.de)).

## METHODS

**Optical attosecond pulse synthesis.** Light waveforms whose spectra extend over more than two optical octaves in the visible and adjacent spectral range (about 1.1–4.6 eV) (Extended Data Fig. 1a) are generated by the nonlinear broadening of laser pulses (approximately 22 fs, 1 mJ, 790 nm) through a hollow-core fibre (HCF) filled with Ne gas (about 2.3 bar). The energy of the generated supercontinuum at the exit of the HCF is about 550  $\mu$ J. The spectra of these pulses are divided by dichroic beam-splitters into four almost equally wide spectral bands centred in the near infrared (NIR; about 1.1–1.75 eV), visible (vis; about 1.75–2.5 eV), visible-ultraviolet (vis-UV; about 2.5–3.5 eV) and deep ultraviolet (DUV; about 3.5–4.6 eV). The pulses in these bands are individually compressed by dispersive mirrors to durations of a few femtoseconds before they are spatially and temporally superimposed to yield a single beam/pulse at the exit of the apparatus. The pulses are temporally characterized by a transient-grating frequency-resolved optical gating (TG-FROG) apparatus. The durations  $T$  of the pulses in different channels of the synthesizer were measured to be  $T_{\text{NIR}} \approx 8.5$  fs,  $T_{\text{vis}} \approx 7$  fs,  $T_{\text{vis-UV}} \approx 6.5$  fs and  $T_{\text{DUV}} \approx 6.5$  fs (see inset to Fig. 2a). The synthesizer apparatus transmits about 82% of the energy of the incoming supercontinuum. As a result, the pulse energy at the exit of the apparatus is about 320  $\mu$ J, and is distributed among the four channels as  $\text{Ch}_{\text{NIR}} \approx 255$   $\mu$ J,  $\text{Ch}_{\text{vis}} \approx 45$   $\mu$ J,  $\text{Ch}_{\text{vis-UV}} \approx 15$   $\mu$ J and  $\text{Ch}_{\text{DUV}} \approx 4$   $\mu$ J, where  $\text{Ch}_i$  denotes the energy of channel  $i$ .

For the synthesis of optical attosecond pulses, both precise control of the relative delay between the constituent pulses in the synthesizer as well as an intensity control over the spectral channels is required. To this end, we followed a new approach, which effectively enables the passive spectral intensity control and facilitates the attosecond streaking characterization of the generated optical waveforms in the same set-up. The EUV attosecond probe is generated first (Extended Data Fig. 1b) by focusing the light transients from the synthesizer into a quasi-static Ne gas cell. The EUV radiation, which emerges collinearly to the driver waveform, is transmitted through a thin, round Zr foil, while the optical pulse, which is transmitted around the geometrical margins of this foil, forms an annular beam. A double-mirror module consisting of a concave, multilayer, coated inner mirror and a metal–dielectric–metal (MDM)-coated concave annular sector (outer mirror) (Extended Data Fig. 2) of the same focal length ( $f = 12.5$  cm), focuses the light transients and the EUV attosecond probe into a second Ne gas nozzle placed near the entrance of a time-of-flight (TOF) spectrometer (Extended Data Fig. 1b). One of the essential characteristics of the MDM is that the imposed spectral control results in negligible phase distortions over the whole spectral range of the supercontinuum pulse. This was experimentally verified by FROG measurements of the pulses in the constituent channels upon reflection off the MDM unit. At the same time, because the EUV probe is generated before the spectral intensity control of the optical transient, it overcomes a fundamental limitation of a half-cycle field: the efficient generation of intense EUV-probe pulses. This is because a half-cycle field does not involve at least two intense field crests, which are required for ionization and subsequent acceleration of electrons<sup>27</sup> leading to high harmonic generation.

Single-cycle optical pulses are generated by physically suppressing a part ( $>3$  eV) of the high-frequency spectrum of the optical attosecond pulse.

**Generation of VUV spectra.** We focus optical attosecond pulses into a quasi-static cell, which replaces the Ne gas jet used for performing attosecond-streaking characterization of their fields, filled with Kr atoms at a moderate pressure (about 80 mbar) (Extended Data Fig. 1b). We probe the nonlinear polarization of the system by recording VUV spectra that emerge collinearly with the driving optical field using a spectrometer placed downstream from the cell. The spectra are sampled at energies higher than about 5 eV, that is, beyond the constituent spectrum of the optical attosecond pulses, and extend no higher than the ionization threshold of neutral Kr ( $I_p \approx 14$  eV).

**Ultimate limit of carrier-envelope phase decomposition.** The concept of the carrier-envelope phase  $\phi_{\text{CE}}$  (CEP) is most readily understood in the time domain<sup>28</sup>. The carrier-envelope phase is the interval  $\Delta t_{\text{peak}}$  between the maximum of the envelope and the maximum of the instantaneous field, translated in phase at the centroid frequency  $\omega_L$  of the spectrum:  $\phi_{\text{CE}} = \omega_L \times \Delta t_{\text{peak}}$ . Any electric field can be decomposed, according to Hilbert's transform as

$$E(t) = A(t) \cos(\omega_L t + \phi_G)$$

in which  $A(t)$  is the envelope (the modulus of the analytical field) and  $\phi_G$  is the global (or absolute) phase.

The maximum of the field is determined by solving

$$\dot{E}(t) = 0 \Leftrightarrow \frac{\dot{A}(t)}{\omega_L A(t)} = \tan(\omega_L t + \phi_G) \quad (2)$$

On the basis of this equation and the estimation  $\frac{\dot{A}(t)}{\omega_L A(t)} \approx \frac{\Delta\omega}{\omega_L}$ , we distinguish two different regimes:

(1) For pulses with durations longer than one cycle,  $\Delta\omega / \omega_L = 1$  and the solution of equation (2) yields  $\phi_G = \omega_L \times \Delta t_{\text{peak}} = \phi_{\text{CE}}$ , which confirms the equivalence between the global ( $\phi_G$ ) and carrier-envelope ( $\phi_{\text{CE}}$ ) phases.

(2) For sub-cycle pulses,  $\Delta\omega / \omega_L > 1$  and so the solution of equation (2) is not trivial because, in a fraction of a cycle, the envelope varies substantially. Consequently,  $\omega_L \times \Delta t_{\text{peak}} < \phi_G$  and the global and carrier-envelope phases are no longer equivalent.

Extended Data Fig. 3a presents the case of a half-cycle pulse—an optical attosecond pulse. The sinusoidal waveform (red) reaches its maximum at 0.2 periods instead of 0.25, which is expected for a wave with a phase of  $\pi/2$  rad. Extended Data Fig. 3b shows the CEP determined using the method of the field maximum depending on the pulse duration, for three settings of the global phase. This comparison reveals a considerable difference between  $\phi_{\text{CE}}$  and  $\phi_G$  in the short pulse regime. In view of the above justification, it is clear that the CEP is an accurate description of the global phase only for pulses longer than one cycle.

**Time-dependent Schrödinger equation and adiabatic models.** To theoretically study the nonlinear dipole dynamics in Kr atoms exposed to intense optical attosecond pulses, we used two models. In the first, we solve the three-dimensional time-dependent Schrödinger equation (TDSE) within the single-active-electron approximation. To this end, we used a central potential for Kr, which was calculated using optimized effective-potential methods<sup>29</sup>. In the second, and in order to describe instantaneous response, we assumed an adiabatic model based on a two-level system<sup>12,30</sup>, in which the dipole moment, in the quasi-static approximation, can be expressed as:

$$\rho(t) = \frac{2d^2 E(t)}{\sqrt{4d^2 [E(t)]^2 + \omega_0^2}}$$

where  $d$  is chosen to match the nonlinear polarizability of Kr (refs 31, 32),  $\omega_0$  is the excitation energy  $\hbar\omega_0 \approx 10$  eV and  $E(t)$  denotes the electric field of the optical attosecond pulse. To access the nonlinear component of the induced electronic dipole moment at a given intensity of the driving field in both models, we perform a second calculation at a much lower (about six orders of magnitude) intensity. As a next step, we subtract the virtually linear dipole calculated at the lower intensity from the original one, after multiplying it by the corresponding ratio between the two intensities. The calculated global-phase spectrograms (spectral emission as a function of the global phase) using the adiabatic and TDSE models are shown in Extended Data Fig. 4a, b. In accordance with the discussion in the main text, the adiabatic model (Extended Data Fig. 4a) predicts uniform modulations of the spectral amplitude of the emitted spectral components as a function of the global phase  $\phi_G$ . In contrast, the spectrogram calculated using the TDSE model embodies the signatures of the delayed electronic response (Extended Data Fig. 4b) in the form of asynchronous amplitude modulations between different frequencies/energies of the emitted dipole. These features are present in the entire emitted spectrum, not only close to the resonant area (10–14 eV). We show that these effects can be used to extract the dynamics of the nonlinear response by reconstructing the global-phase ( $\phi_G$ ) spectrograms recorded in our experiments.

Extended Data Fig. 4c shows a global-phase spectrogram simulated for a single-cycle pulse using the TDSE Kr model. In this regime of single-cycle pulses, the global-phase spectrogram does not show discernible variation over  $\phi_G$ , which is experimentally verified (see Fig. 3e–g).

**Ionization-free strong field polarization of bound electrons.** An essential innovation introduced by using optical attosecond pulses is their unique capability to drive nonlinear dynamics in quantum systems without inducing a substantial degree of ionization or excitation, that is, without greatly altering the original system. In experiments where the polarization of the system is to be probed, both excessive ionization and excitations markedly modify the system, resulting in a considerable degree of ‘contamination’ in the emitted signal from the new atomic entities; such contamination is challenging to resolve both experimentally and theoretically.

This capability is unique to isolated sub-cycle pulse structures and is not observed for trains of sub-cycle field modulation<sup>33</sup> because the nanosecond-long exposure of atoms to such fields yields a substantial degree of ionization, which is actually the means to trace the waveform in these experiments<sup>34</sup>. The ionization and excitation probability of Kr atoms calculated by the solution of the TDSE and the waveform used in our experiments for a range of intensities (summarized in Extended Data Table 1) verify this conjecture. Experimentally, we also verified this fact by using the previously established and highly sensitive technique of attosecond transient absorption spectroscopy<sup>18</sup>.



The interaction of optical attosecond pulses with matter, even at excessive intensities such as those used in our experiments and simulations, can be mostly considered as a scattering process. The system and the pulse virtually do not exchange energy; rather, the pulse probes the system via nonlinear scattering and the emission of coherent radiation.

**Nonlinearities in the interaction of Kr with intense optical attosecond pulses.** Identifying the dominant nonlinearities in the interaction between the optical attosecond pulses and the Kr atoms, and understanding the dependence of these nonlinearities on the variation of the global phase ( $\phi_G$ ) of the optical attosecond pulse—the key control point of the interaction in experiments with such waveforms—is essential for the development of intuitive models that can describe the nonlinear dynamics. As we show below, it allows the development of a robust methodology that permits the reconstruction of the dynamics of the response from global-phase spectrograms.

To this end, we theoretically (using TDSE simulations in Kr) studied the intensity dependence of the yield of the spectral emission (averaged over the range 0–8 eV) under optical attosecond pulses for the range of intensity settings used in our experiments. The corresponding nonlinearities can be evaluated from the slope of the linear fit of the data in the log–log diagram of Extended Data Fig. 5. A linear fitting over the entire range of intensities studied ( $2 \times 10^{13}$ – $10 \times 10^{13}$  W cm $^{-2}$ ) reveals a slope of about 4, and suggests the dominance and coexistence of the two most essential nonlinearities in centrosymmetric systems: the third- ( $E^3$ ) and the fifth- ( $E^5$ ) order nonlinearities—broadly known as bound-electronic nonlinearities.

**Simple model of the field-driven bound-electron nonlinear response of Kr.** The study presented in the previous paragraph highlights the dominance of bound-electron nonlinearities in the response and the low or negligible sensitivity of the nonlinearity to the global phase of our pulses.

To account for the non-instantaneous response revealed in our TDSE simulations, and inspired by previous approaches in ultrafast spectroscopy<sup>35</sup>, in which the nonlinear response is decomposed into instantaneous and delayed components, we describe the nonlinear dipole moment as a sum over instantaneous third- and fifth-order nonlinearities, as well as the fifth-order delayed nonlinearity according to equation (1):

$$\rho(t, \phi_G) = aE^3(t, \phi_G) + bE^5(t, \phi_G) + cE^5(t - dt, \phi_G)$$

Here,  $a$ ,  $b$  and  $c$  are coefficients,  $dt$  represents a delay of the fifth-order response and  $E(t, \phi_G)$  is the electric field of the optical attosecond pulse for a global phase  $\phi_G$ .

One would generally expect delayed terms to be considered for all nonlinearities involved (including the third-order); the energy diagram of Extended Data Fig. 6 offers an intuitive explanation of our choice to limit the delayed terms to only fifth-order nonlinearities. Indeed, delayed response in the range 0–9 eV will primarily involve virtual transitions, which can coherently couple at least two electronic states of the system (ground and excited states or combinations of excited states). The diagram demonstrates this assuming virtual transition compatible with the energy spectrum of our optical attosecond pulse (1.1–4.6 eV). Such transitions can only occur within the fifth-order response (higher-order Kerr effects). The diagram also highlights—through the multiphoton picture—that at this extreme limit of pulse duration and corresponding bandwidth, non-resonant and resonant response are virtually inseparable. As a result, the coherent dynamics induced between two or more states of the system will be manifested at each nonlinearly emitted spectral component in the process.

Extended Data Fig. 7a–c shows representative, synthetic, global-phase spectrograms for three values of the parameter  $dt$  ( $dt = 0$  (instantaneous response),  $dt = 20$  as and  $dt = 30$  as), generated by equation (1) in the spectral range of our experiments (see Fig. 4). The synthetic spectrograms highlight the capability of the model to capture key features of the experimental spectrograms (for example, Figs 1 or 4), such as the profoundly asynchronous modulation of the emission in the range 7–8 eV and the weakening of the amplitude in the range 6–6.5 eV—both are the result of dynamic nonlinear interference between delayed and instantaneous terms in equation (1).

To further verify the validity of our model, we used the TDSE simulations in Kr as a basis to explore how a dipole described by the above equations could represent the (below resonances) nonlinear response of the system.

Extended Data Fig. 7d shows the TDSE simulated dipoles (black line) and their fitting with equation (1). We investigated both the capability of the model to fit only a fraction (spectrally filtered from 5.5–8 eV; Extended Data Fig. 7d) of

the dipole as well the entire dipole (0–8 eV) below resonances (Extended Data Fig. 7e). A single set of the parameters  $a$ ,  $b$ ,  $c$  and  $dt$  in equation (1), and a given peak intensity of the pulse, can precisely reproduce the nonlinear dipole (black lines) for all settings of the global phase. Representative examples for two (extreme) settings of the global phase  $\phi_G \approx 0$  (left panels of Extended Data Fig. 7d and e) and  $\phi_G \approx \pi/2$  rad (right panels of Extended Data Fig. 7d and e) are shown. Furthermore, the parameters  $a$ ,  $b$ ,  $c$  and  $dt$  extracted from fitting part of the nonlinear spectrum (5.5–8 eV) (Extended Data Fig. 7d) are identical to those required to fit the entire range (0–8 eV) (Extended Data Fig. 7e). As a conclusion, even a limited fraction of the nonlinear spectrum contains information about the interfering terms in equation (1). The model works well for a wide range of peak intensities (not shown); therefore, we conclude that it is adequate for reconstructing the response.

**Nonlinear dipole reconstructions.** The most critical test of the capability of our model to reconstruct experimental data and to trace the time-domain nonlinear dipole dynamics, is to perform a numerical experiment. To do so, we create a theoretical spectrogram using the TDSE model (where the nonlinear dipole is known *a priori*) and attempt to reconstruct this nonlinear dipole using (i) equation (1), (ii) the spectrogram constructed by the TDSE simulation, and (iii) the *a priori* measured driver field waveform.

For both experimental and synthetic data, we used a quickly converging nonlinear algorithm to perform the reconstructions. It is based on a commercial, highly optimized numerical routine that uses the 'trust-region' method<sup>36</sup>, which is usually used for constrained problems. The root-mean-square deviation (r.m.s.d.) was determined, and defined as the main parameter to optimize our reconstruction via its minimization. The r.m.s.d. is calculated as:

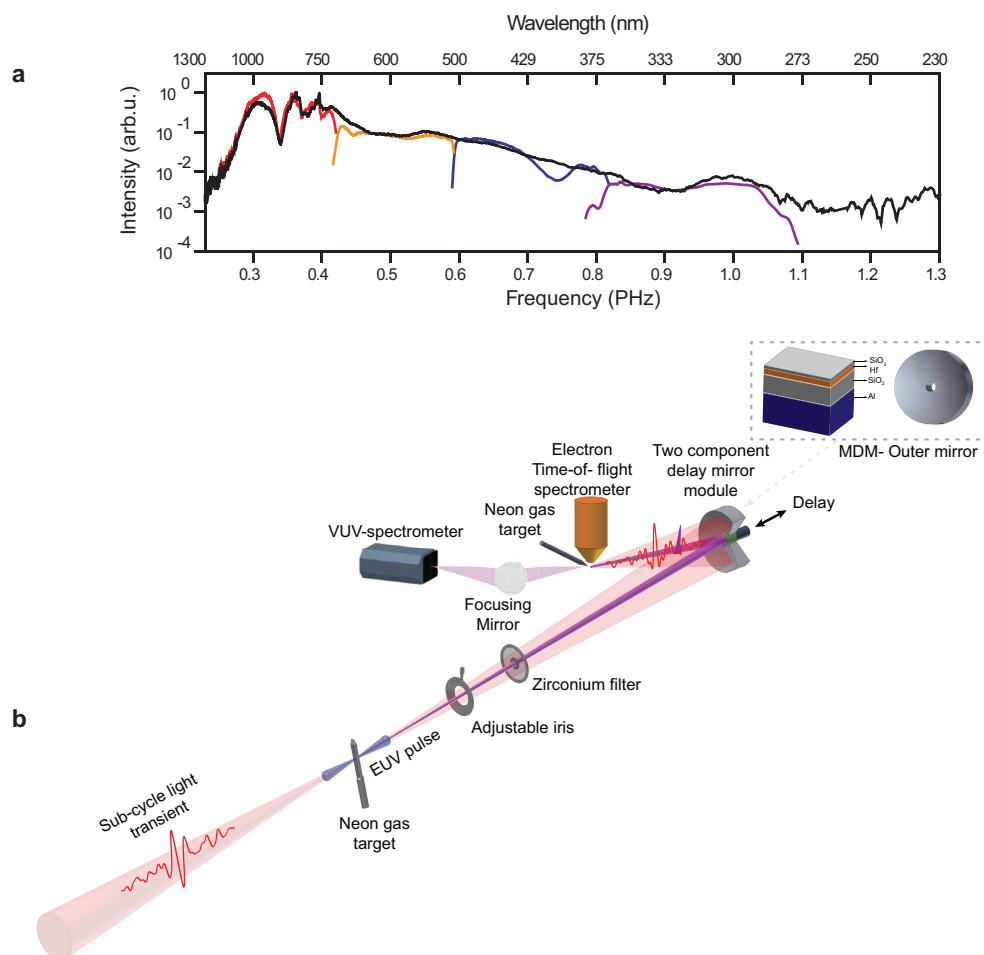
$$\text{r.m.s.d.} = \sqrt{\frac{\sum_{i=1}^{n_{\phi_G}} \int_{\omega_1}^{\omega_2} (\hat{X}_i - X_i)^2 d\omega}{n_{\phi_G} \times n_{\omega}}}$$

Where  $\hat{X}$  and  $X$  are the computed and original (measured) spectra, respectively, at certain global-phase settings,  $\omega_1 = 5.5$  eV,  $\omega_2 = 8$  eV,  $n_{\phi_G}$  is the number of global-phase settings of the measured VUV spectrogram, and  $n_{\omega}$  is the number of frequencies involved (from  $\omega_1$  to  $\omega_2$ ).

The results of this study are summarized in (Extended Data Fig. 8a, b), in which the original and the reconstructed spectrograms are shown. The reconstruction parameters in equation (1) are  $a = 0.180$ ,  $b = 0.096$ ,  $c = -0.177$  and  $dt = 67$  as. Extended Data Figure 8c, d compares the reconstructed (red line) and the *a priori* known nonlinear dipole (black line) for two different settings of the global phase. Blue lines show the instantaneous response for the same waveforms for comparison.

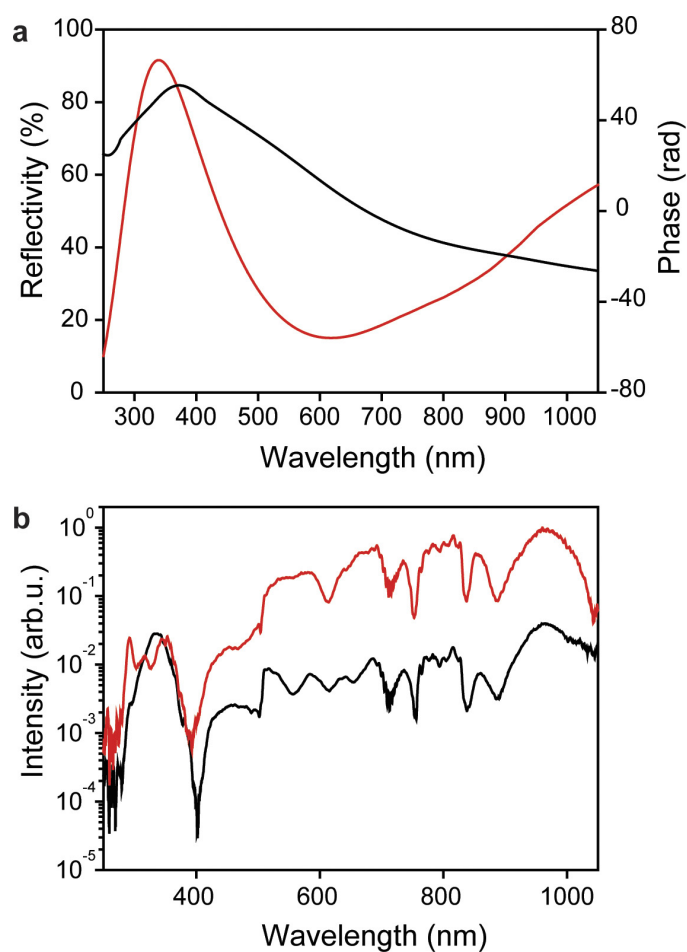
**Intensity dependent nonlinear delay.** We study the delayed nonlinear bound-electronic response over a wide range of intensities ( $(2\text{--}8) \times 10^{13}$  W cm $^{-2}$ ). The evaluated delays between the instantaneous and the TDSE-simulated dipoles as function of the driver field intensity are shown in Extended Data Fig. 9a; the delays between the instantaneous dipoles and those reconstructed from our measured spectrograms (see Fig. 4) are shown in Extended Data Fig. 9b.

27. Corkum, P. B. Plasma perspective on strong field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
28. Krausz, F. & Ivanov, M. Attosecond physics. *Rev. Mod. Phys.* **81**, 163–234 (2009).
29. Muller, H. G. An efficient propagation scheme for the time-dependent Schrödinger equation in the velocity gauge. *Laser Phys.* **9**, 138–148 (1999).
30. Gauthier, F. I., Garraway, B. M. & Knight, P. L. High harmonic generation and periodic level crossings. *Phys. Rev. A* **56**, 3093–3096 (1997).
31. Lundeen, T., Hou, S. Y. & Nibler, J. W. Nonresonant third order susceptibilities for various gases. *J. Chem. Phys.* **79**, 6301–6305 (1983).
32. Shelton, D. P. Nonlinear-optical susceptibilities of gases measured at 1064 and 1319 nm. *Phys. Rev. A* **42**, 2578–2592 (1990).
33. Harris, S. E. & Sokolov, A. V. Subfemtosecond pulse generation by molecular modulation. *Phys. Rev. Lett.* **81**, 2894–2897 (1998).
34. Baker, S., Walmsley, I. A., Tisch, J. W. G. & Marangos, J. P. Femtosecond to attosecond light pulses from a molecular modulator. *Nature Photon.* **5**, 664–671 (2011).
35. Blow, K. J. & Wood, D. Theoretical description of transient stimulated Raman scattering in optical fibers. *IEEE J. Quantum Electron.* **25**, 2665–2673 (1989).
36. Coleman, T. F. & Li, Y. Y. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM J. Optim.* **6**, 418–445 (1996).



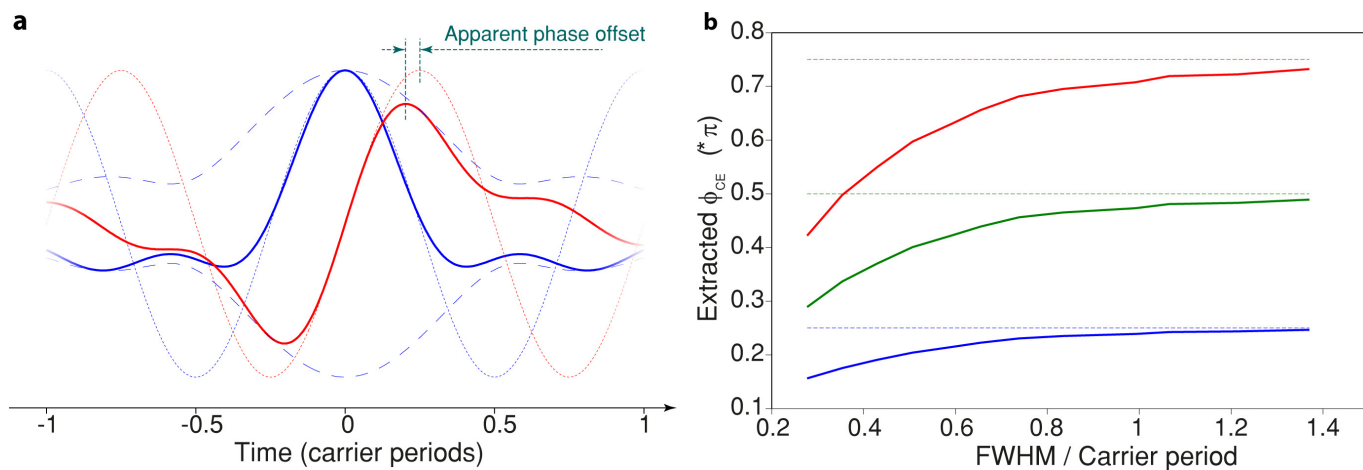
**Extended Data Figure 1 | Supercontinuum spectra and experimental set-up.** **a**, The supercontinuum spectra spans more than two octaves (1.1–4.6 eV), and is shown by the black line; the spectra of the individual

channels  $\text{Ch}_{\text{NIR}}$ ,  $\text{Ch}_{\text{vis}}$ ,  $\text{Ch}_{\text{vis-UV}}$  and  $\text{Ch}_{\text{DUV}}$  are shown by the red, orange, blue and violet lines, respectively. **b**, Schematic illustration of the experiment set-up.



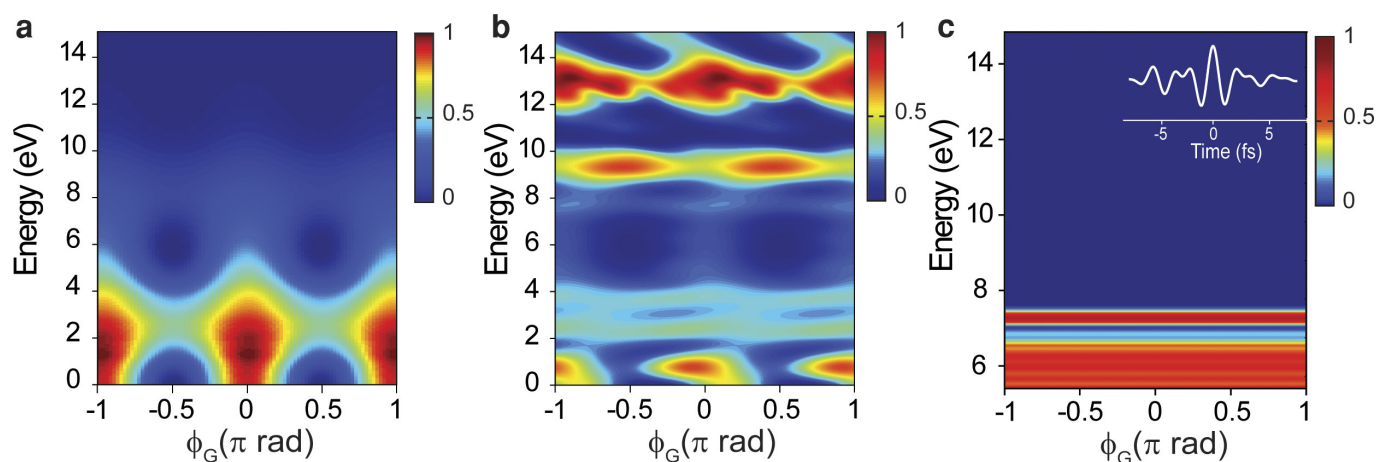
**Extended Data Figure 2 | Spectral equalization by a metal-dielectric-metal (MDM) coating.** **a**, The designed reflectivity (red line) and spectral phase (black) of the MDM coating applied on the outer mirror of the double mirror module shown in Extended Data Fig. 1. **b**, Multi-octave spectrum at the exit of the synthesizer (red line) and upon reflection off the MDM mirror (black line).





**Extended Data Figure 3 | Global phase and CEP difference in a sub-cycle regime.** **a**, Half-cycle, cosine-like (blue line) and sine-like (red line) waveforms, compared with the carrier wave (dotted blue and red lines) and envelope (dashed blue lines). **b**, Limits of CEP decomposition in the time

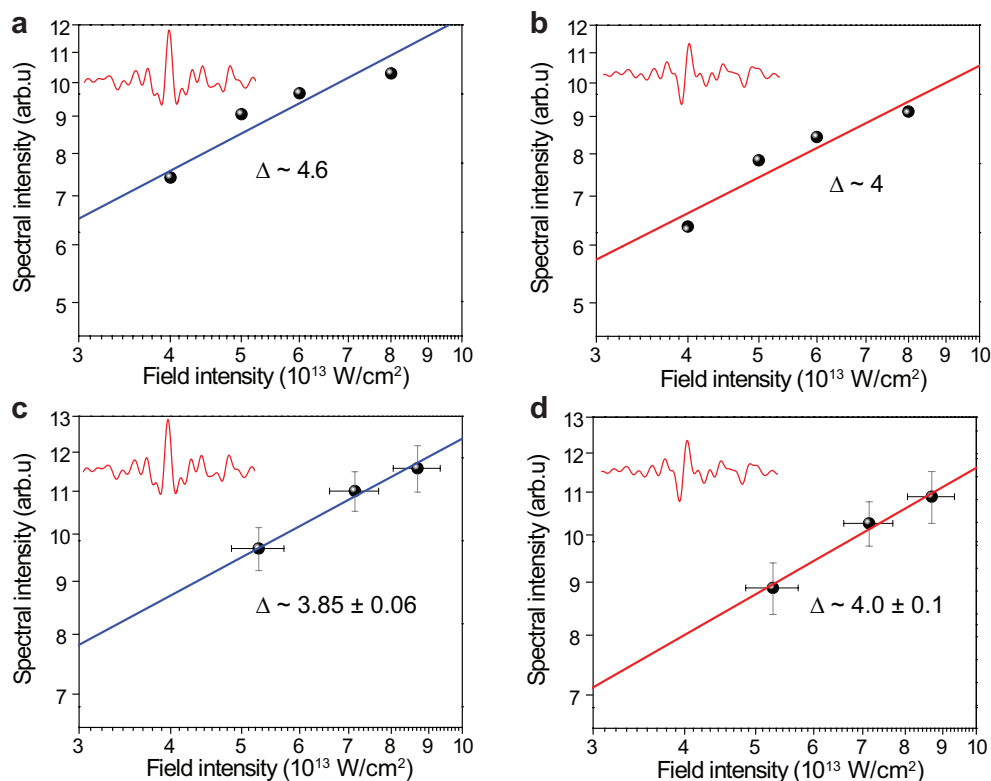
domain. Retrieved CEP (solid lines) versus global-phase offset (dashed lines) as a function of the ratio between the FWHM of the intensity envelope and the period of the carrier wave for  $\phi_G = \pi/4$  rad (blue line),  $\phi_G = \pi/2$  rad (green line) and  $\phi_G = 3\pi/4$  rad (red).



**Extended Data Figure 4 | Simulated global-phase spectrograms.**

**a, b,** Calculated global-phase ( $\phi_G$ ) spectrograms using the adiabatic model (**a**) and the TDSE model (**b**) for experimentally sampled waveforms (optical attosecond pulse) and an intensity of about  $4 \times 10^{13} \text{ W cm}^{-2}$ .

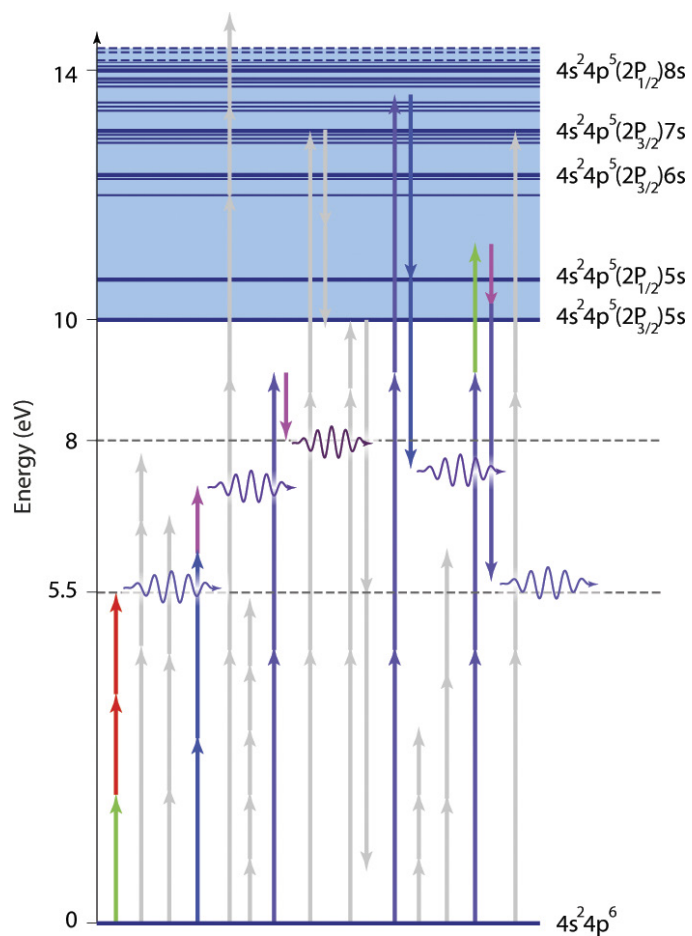
**c,** Simulated global-phase spectrograms for a sub-cycle pulse (about 2 fs; shown in inset). The polarization response is insensitive to global-phase variation of the waveform; see also experiments in Fig. 3. The colour scale represents the spectral intensity in arbitrary units.



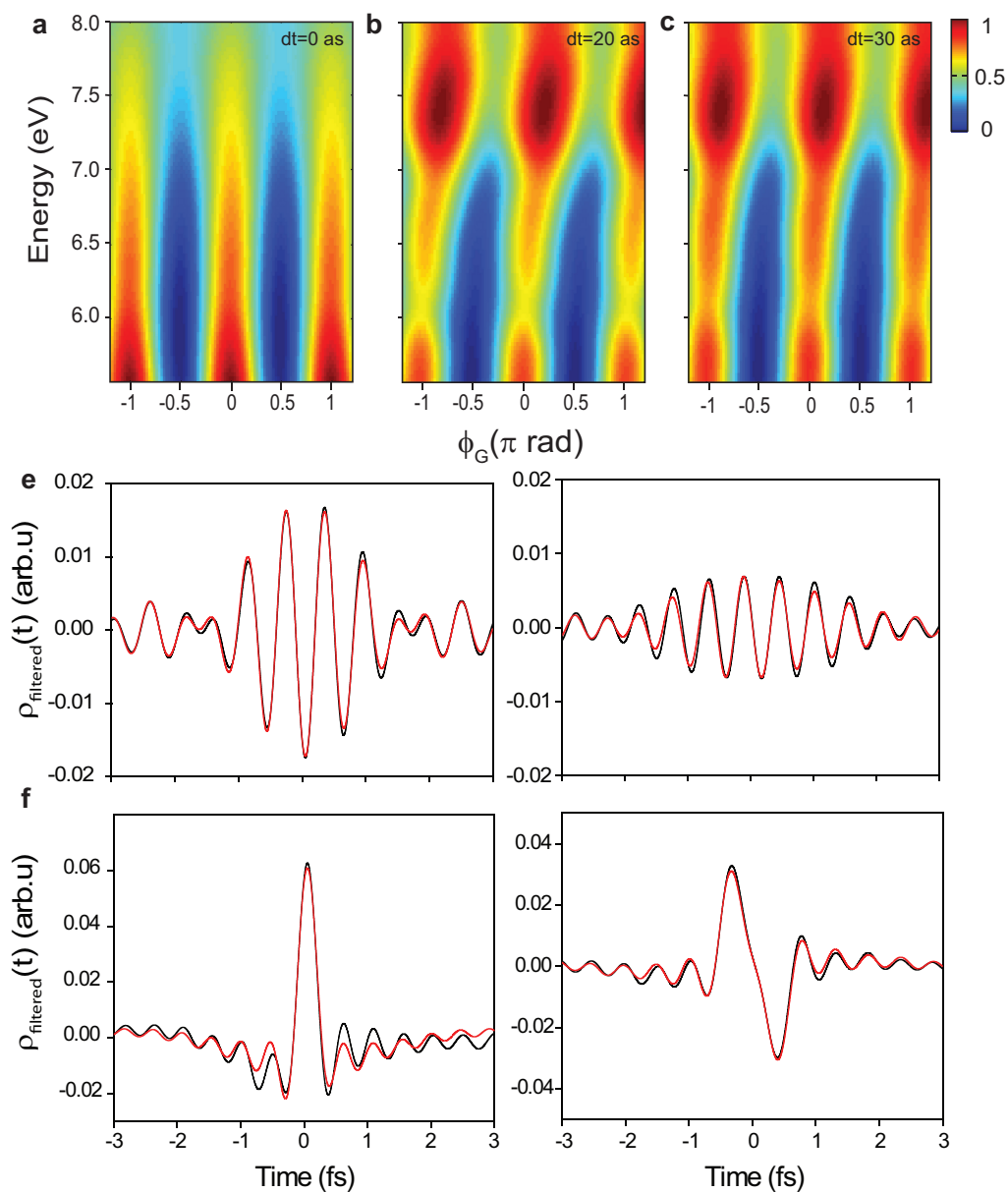
**Extended Data Figure 5 | Nonlinearity study of the interaction of an optical attosecond pulse with a Kr atom. a–d,** Data points show the spectrally integrated yield of the nonlinear polarization (0–8 eV) as a function of the intensity of the optical attosecond pulse for two global phases  $\phi_G$  of the driving waveform,  $\phi_G \approx 0$  (**a, c**) and  $\phi_G \approx \pi/2$  rad (**b, d**),

evaluated from TDSE simulations (**a, b**) and our experiments in Fig. 4 (**c, d**). The error bars represent the standard error of the mean of the input field intensity and the calculated integrated spectral yield of the nonlinear polarization.  $\Delta$  is the slope of the linear fitting of the log–log plots; the errors shown indicate the standard error of mean, with  $n = 3$ . Insets show



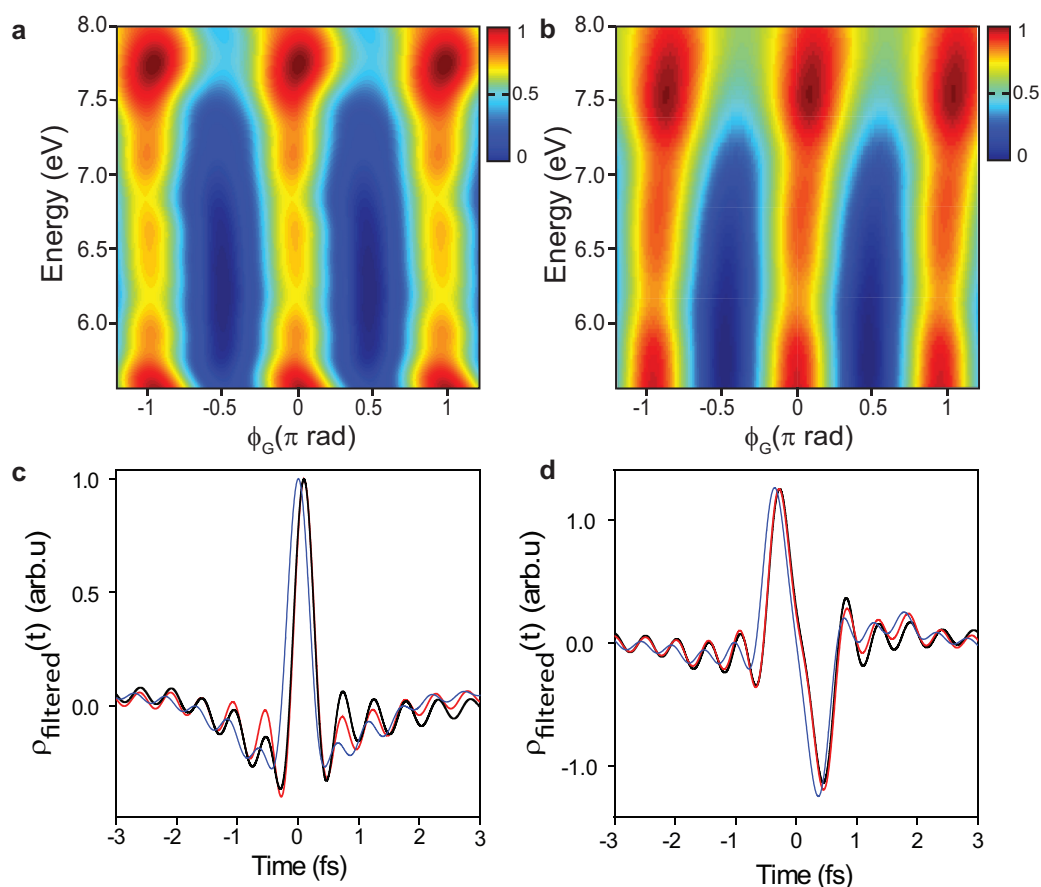


Extended Data Figure 6 | Representation of the possible multiphoton transitions on a Kr atom energy level diagram assuming third- and fifth-order nonlinear processes.



**Extended Data Figure 7 | Synthetic global-phase spectrograms and nonlinear dipole fittings.** a–c, Synthetic global-phase ( $\phi_G$ ) spectrograms determined using the simple model (equation (1)) with the same set of parameters ( $a$ ,  $b$ ,  $c$ ) and delays  $dt = 0$  as (a),  $dt = 20$  as (b) and  $dt = 30$  as (c).

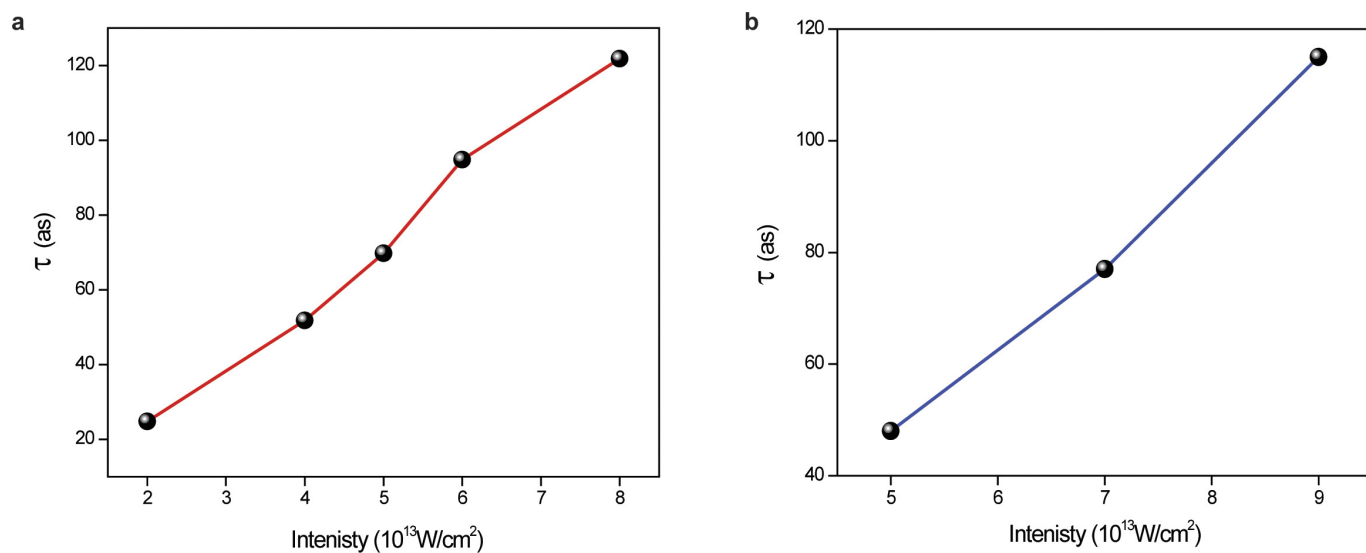
The colour scale represents spectral intensity in arbitrary units. d, e, Fits (red lines) of the spectrally filtered (d, 5.5–8 eV; e, 0–8 eV) nonlinear dipoles ( $\rho(t)$ , black lines) with  $\phi_G \approx 0$  (left column) and  $\phi_G \approx \pi/2$  rad (right column) using equation (1) and  $6 \times 10^{13} \text{ W cm}^{-2}$ .



**Extended Data Figure 8 | Reconstruction of the global-phase spectrogram calculated by the TDSE model.** **a**, Calculated spectrogram at different global-phase settings ( $\phi_G = -1.2\pi - 1.2\pi$  rad) of the optical attosecond pulse using the TDSE model. **b**, Reconstructed spectrogram using the model of equation (1). The colour bar represents spectral

intensity in arbitrary units. **c**, **d**, Calculated nonlinear dipoles using the TDSE model at  $\phi_G \approx 0$  (**c**) and  $\phi_G \approx \pi/2$  rad (**d**) are plotted in black; the reconstructed nonlinear dipoles and the adiabatic dipoles are shown in red and blue, respectively.





**Extended Data Figure 9 | Intensity dependence of the nonlinear dipole delay.** a, b, The delay between the adiabatic dipole and the calculated dipole using the TDSE model as a function of the peak intensity according to TDSE simulations (a) and evaluated from the experimental data (b; see Fig. 4).

**Extended Data Table 1 | Ionization and excitation probabilities calculated using TDSE simulations of Kr for a range of peak intensities of the optical attosecond pulse**

Peak intensity ( $10^{13}$ W/cm <sup>2</sup> )	1	2	4	6	8	10
Ionization probability (%)	0.0018	0.06	0.18	0.84	1.38	2.4
Excitation probability ( $10^{-4}$ %)	0.004	0.05	0.825	1.88	3.63	3.8

# Bioresorbable silicon electronic sensors for the brain

Seung-Kyun Kang<sup>1,2\*</sup>, Rory K. J. Murphy<sup>3\*</sup>, Suk-Won Hwang<sup>4\*</sup>, Seung Min Lee<sup>1,2\*</sup>, Daniel V. Harburg<sup>1,2</sup>, Neil A. Krueger<sup>1</sup>, Jiho Shin<sup>2,5</sup>, Paul Gamble<sup>3</sup>, Huanyu Cheng<sup>6</sup>, Sooyoun Yu<sup>2,5</sup>, Zhuangjian Liu<sup>7</sup>, Jordan G. McCall<sup>8</sup>, Manu Stephen<sup>3</sup>, Hanze Ying<sup>1</sup>, Jeonghyun Kim<sup>1,2</sup>, Gayoung Park<sup>9,10</sup>, R. Chad Webb<sup>1,2</sup>, Chi Hwan Lee<sup>11</sup>, Sangjin Chung<sup>1,2</sup>, Dae Seung Wie<sup>12</sup>, Amit D. Gujar<sup>3</sup>, Bharat Vemulapalli<sup>3</sup>, Albert H. Kim<sup>3</sup>, Kyung-Mi Lee<sup>10</sup>, Jianjun Cheng<sup>1</sup>, Younggang Huang<sup>13</sup>, Sang Hoon Lee<sup>14</sup>, Paul V. Braun<sup>1,2,15</sup>, Wilson Z. Ray<sup>3</sup> & John A. Rogers<sup>1,2,15</sup>

Many procedures in modern clinical medicine rely on the use of electronic implants in treating conditions that range from acute coronary events to traumatic injury<sup>1,2</sup>. However, standard permanent electronic hardware acts as a nidus for infection: bacteria form biofilms along percutaneous wires, or seed haematogenously, with the potential to migrate within the body and to provoke immune-mediated pathological tissue reactions<sup>3,4</sup>. The associated surgical retrieval procedures, meanwhile, subject patients to the distress associated with re-operation and expose them to additional complications<sup>5–8</sup>. Here, we report materials, device architectures, integration strategies, and *in vivo* demonstrations in rats of implantable, multifunctional silicon sensors for the brain, for which all of the constituent materials naturally resorb via hydrolysis and/or metabolic action<sup>9–12</sup>, eliminating the need for extraction. Continuous monitoring of intracranial pressure and temperature illustrates functionality essential to the treatment of traumatic brain injury<sup>2,13</sup>; the measurement performance of our resorbable devices compares favourably with that of non-resorbable clinical standards. In our experiments, insulated percutaneous wires connect to an externally mounted, miniaturized wireless potentiostat for data transmission. In a separate set-up, we connect a sensor to an implanted (but only partially resorbable) data-communication system, proving the principle that there is no need for any percutaneous wiring. The devices can be adapted to sense fluid flow, motion, pH or thermal characteristics, in formats that are compatible with the body's abdomen and extremities, as well as the deep brain, suggesting that the sensors might meet many needs in clinical medicine.

Figure 1a and Supplementary Fig. 1 show a bioresorbable pressure sensor with a magnified illustration of the active region and its cross-sectional side view. The construction involves a membrane of poly(lactic-co-glycolic acid) (PLGA, with a thickness of 30  $\mu\text{m}$ ), sealed against a supporting substrate of nanoporous silicon (60–80  $\mu\text{m}$  thick; 71% porosity) or magnesium foil (60–80  $\mu\text{m}$  thick; see Supplementary Figs 2, 3). The substrate has a square structure of relief (with a depth of 30–40  $\mu\text{m}$ ) etched onto its surface. The associated air cavity allows the membrane to deflect in response to pressure in the fluid surroundings. A silicon nanomembrane in a serpentine geometry serves as a piezoresistive element that rests on the surface

of the membrane near one of the edges of the cavity, where deflection-induced strains are largest (Fig. 1b). The resistance of this sensing element increases monotonically in a linear fashion across the full range of pressures that are relevant to intracranial monitoring (that is, 0–70 mm Hg). An overcoat of silicon oxide ( $\text{SiO}_2$ , about 100 nm thick) provides electrical passivation and a barrier against biofluids. Figure 1c and Supplementary Fig. 4 show photographs of two representative devices of different dimensions to illustrate the scalability of fabrication; the total sizes and weights are 1 mm  $\times$  2 mm  $\times$  0.08 mm (trench size: 0.67 mm  $\times$  0.8 mm  $\times$  0.03 mm) and about 0.4 mg; and 3 mm  $\times$  6 mm  $\times$  0.11 mm (trench size: 2 mm  $\times$  2.4 mm  $\times$  0.04 mm) and roughly 1 mg, respectively.

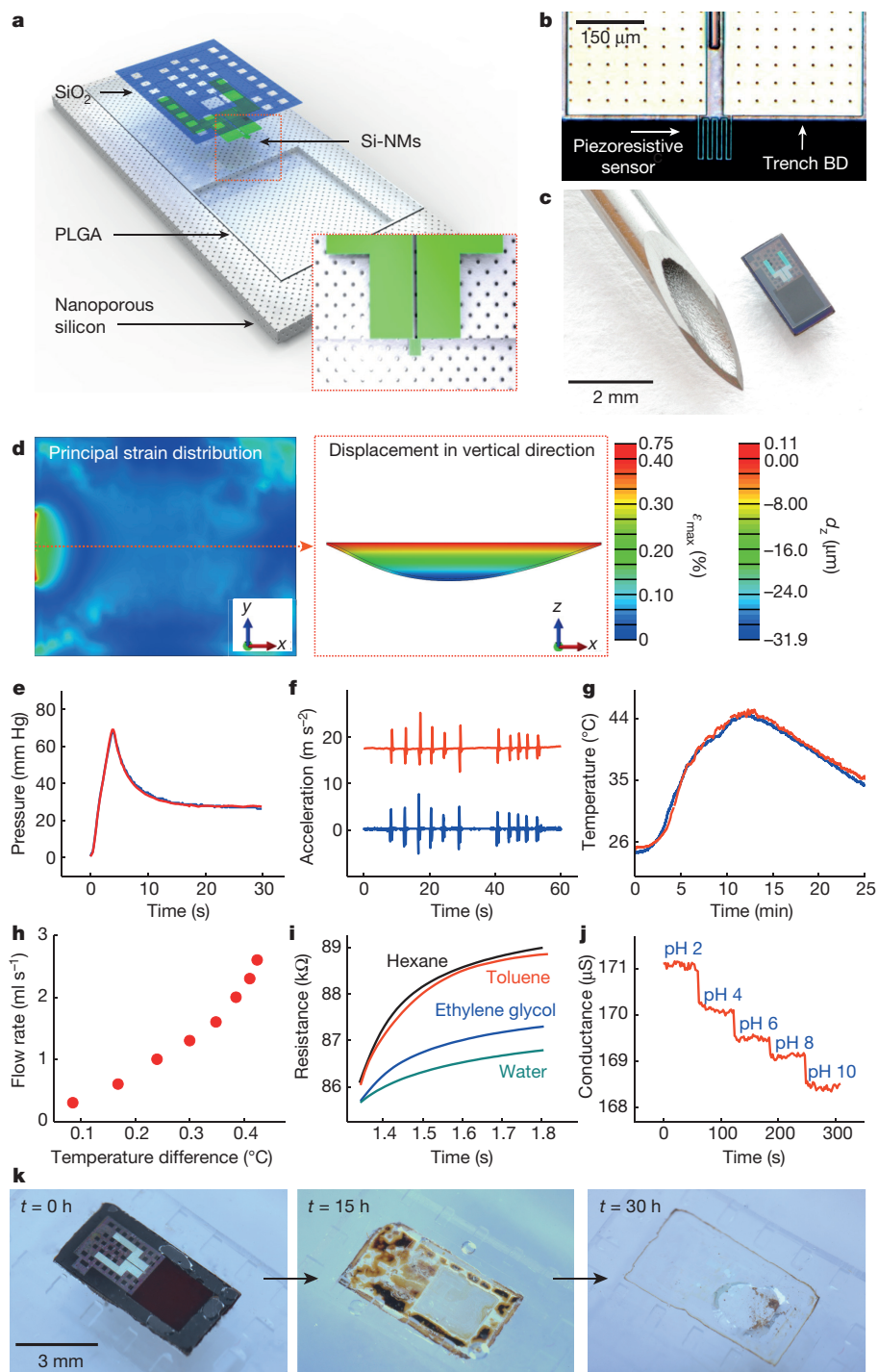
The mechanics of the system can be captured quantitatively by three-dimensional finite element analysis (FEA). Distributions of principal strains and vertical displacements evaluated at an external pressure of 50 mm Hg appear in Fig. 1d. The maximum strain for any applied pressure over the range of interest occurs at the midpoint of the left (and right) edge of the trench, thus motivating this choice of location for the silicon-nanomembrane piezoresistive element (see Supplementary Methods and Supplementary Figs 5 and 6 for details)<sup>14</sup>. The calibration between pressure and resistance is linear, with a slope of 83  $\Omega(\text{mm Hg})^{-1}$ , consistent with modelling results and a gauge factor of about 30, which lies within a range of expected values for monocrystalline silicon (Supplementary Fig. 7)<sup>15</sup>.

Evaluations in set-ups that resemble the intracranial cavity reveal measured pressure responses that agree quantitatively with those of clinical-standard, non-bioresorbable sensors (Fig. 1e and Supplementary Figs 8–10). With various simple modifications, this same platform can be used for precision measurement of other parameters of interest in biomedicine and clinical care. Examples include: motion sensors built with a cantilevered test mass of PLGA (that is, a single-axis accelerometer, Fig. 1f); temperature sensors that exploit the temperature-dependent resistance of silicon-nanomembrane elements set apart from the cavity structure (Fig. 1g); flow sensors in which the silicon nanomembranes serve simultaneously as heating elements and temperature sensors (Fig. 1h); thermal conductivity/diffusivity sensors that exploit related concepts (Fig. 1i); and pH sensors that rely on electrostatic gating of transport through the silicon nanomembrane (Fig. 1j). In addition, chemically functionalizing the surface of the silicon of this last device provides a route to biomolecular sensing, using

<sup>1</sup>Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. <sup>2</sup>Frederick Seitz Materials Research Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. <sup>3</sup>Department of Neurological Surgery, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>4</sup>KU-KIST Graduate School of Converging Science and Technology, Korea University, Seoul 136-701, Republic of Korea. <sup>5</sup>Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. <sup>6</sup>Department of Engineering Science and Mechanics, Materials Research Institute, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>7</sup>Institute of High Performance Computing, Singapore 138632, Singapore. <sup>8</sup>Department of Anesthesiology, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>9</sup>Department of Biomicrosystem Technology, Korea University, Seoul 136-701, South Korea. <sup>10</sup>Department of Biochemistry and Molecular Biology, Korea University College of Medicine, Seoul 136-713, South Korea. <sup>11</sup>Weldon School of Biomedical Engineering, School of Mechanical Engineering, The Center for Implantable Devices, Birck Nanotechnology Center, Purdue University, West Lafayette, Indiana 47907, USA. <sup>12</sup>School of Mechanical Engineering, Purdue University, West Lafayette, Indiana 47907, USA. <sup>13</sup>Department of Mechanical Engineering, Civil and Environmental Engineering, Materials Science and Engineering, and Skin Disease Research Center, Northwestern University, Evanston, Illinois 60208, USA. <sup>14</sup>Department of Biomedical Engineering, College of Health Science, Korea University, Seoul 136-703, South Korea. <sup>15</sup>Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

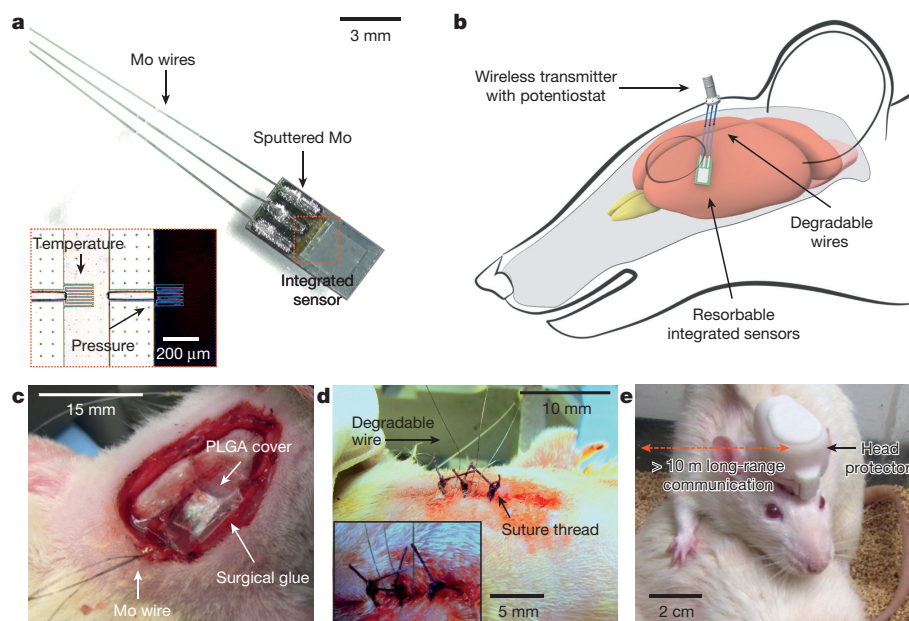
\*These authors contributed equally to this work.





**Figure 1 | Bioresorbable, silicon-based mechanical/physical/chemical sensors for biomedical applications.** **a**, Schematic illustration of a biodegradable pressure sensor. The inset shows the location of the silicon-nanomembrane (Si-NM) strain gauge. **b**, Optical micrograph of the strain-gauge region. ‘Trench BD’, boundary of the trench. **c**, Image of a complete device. The outer diameter of the hypodermic needle is 1 mm. **d**, Left, distribution of principal strains across the PLGA layer, including the Si-NM strain gauge at the left edge, determined from finite element analysis (FEA) for an external pressure of 50 mm Hg. Right, corresponding displacement profile evaluated along the red dotted line in the left panel.  $\epsilon_{\max}$  and  $d_z$  are the principal strain and vertical displacement, respectively. **e**, Responses of a commercial pressure sensor (blue) and a calibrated biodegradable device (red) to time-varying pressure over a range relevant to intracranial monitoring. **f**, Response of a similar biodegradable device (red), but configured as an accelerometer, with comparison to a commercial sensor (blue). **g**, Comparison of the calibrated response of

such a bioresorbable temperature sensor (red) to a commercial device (blue). **h**, The difference in temperature measured by two separate Si-NM temperature sensors placed near a Si-NM element for Joule heating allows assessment of flow rate. **i**, A single serpentine Si-NM used as both a temperature sensor and a heating element allows measurements of thermal conductivity and heat capacity. The graph shows time-dependent changes in temperature upon actuation of Joule heating in devices immersed in different liquids. The coefficients of thermal conductivity ( $\kappa$ , measured through the rate of resistance change) of hexane, toluene, ethylene glycol, and water are 0.12, 0.13, 0.26, and 0.60 W m<sup>-1</sup> K<sup>-1</sup>, respectively. **j**, When the Si-NM is exposed to aqueous surroundings, its conductance depends on pH. The graph shows measurements for immersion in solutions with pH values between 2 and 10. **k**, Images collected at several stages of accelerated dissolution of a bioresorbable pressure sensor upon insertion into an aqueous buffer solution (pH 12) in a transparent PDMS enclosure at room temperature.



**Figure 2 | Bioresorbable interfaces between intracranial sensors and external wireless data-communication modules with percutaneous wiring.** **a**, Image of bioresorbable pressure and temperature sensors integrated with dissolvable metal interconnects (sputtered molybdenum, Mo, 2  $\mu\text{m}$  thick) and wires (Mo, 10  $\mu\text{m}$  thick). The inset shows an optical micrograph of the serpentine Si-NM structures that form the sensing regions. The Si-NM that is not above the air cavity (left) responds only to temperature; the one at the edge of the air cavity (right) responds primarily

to pressure. **b**, Diagram of a bioresorbable sensor system in the intracranial space of a rat, with electrical interconnects that provide an interface to an external wireless data-transmission unit for long-range operation.

**c**, **d**, Demonstrations of **c**, an implanted bioresorbable sensor in a rat, and **d**, a sutured individual. A thin film of PLGA ( $\sim 80 \mu\text{m}$ ) and a degradable surgical glue (TISSEAL) seal the craniectomy defect to close the intracranial cavity. **e**, Healthy, freely moving rat equipped with a complete, biodegradable wireless intracranial sensor system.

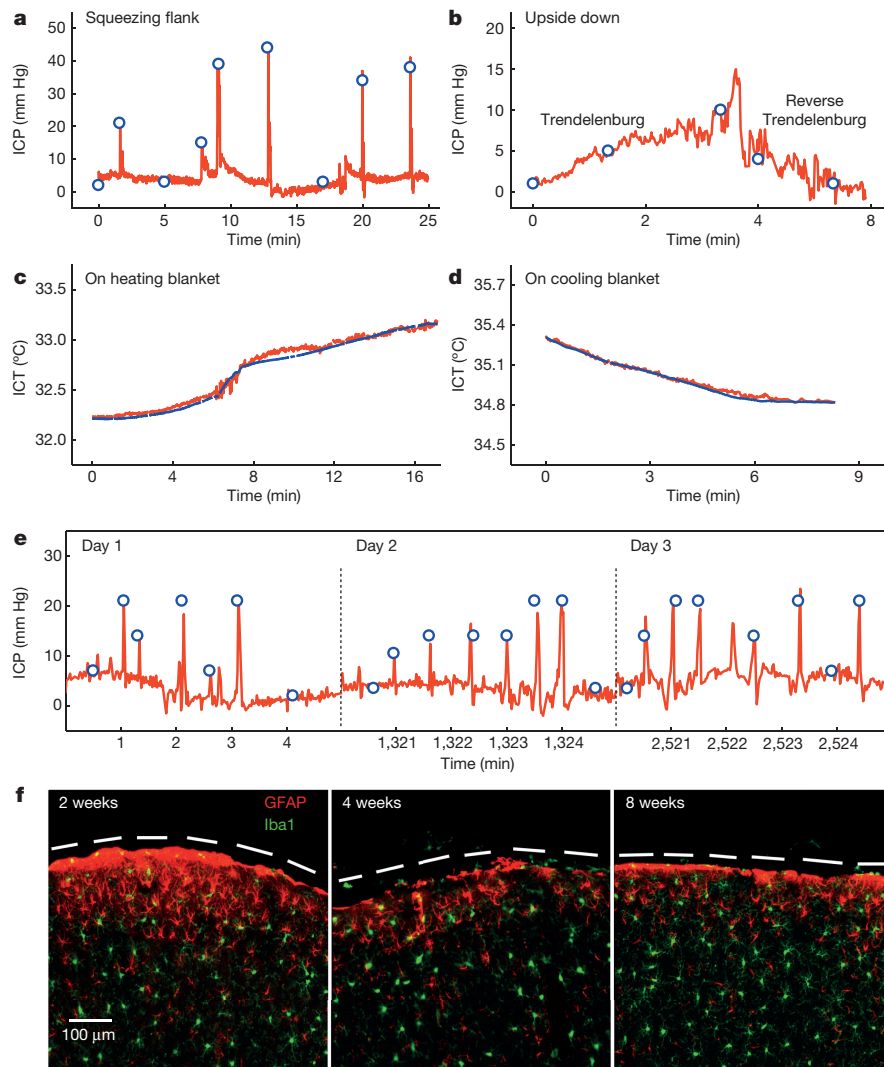
schemes similar to those in conventional silicon biosensors<sup>16–18</sup>. The fabrication methods and operating principles for each of the modalities in Fig. 1f–j appear in Supplementary Methods and Supplementary Figs 11–15.

The uniqueness of these devices is their ability to dissolve completely into biocompatible end products when immersed in aqueous solutions, including biofluids such as cerebrospinal fluid (CSF). Hydrolysis of the silicon nanomembranes, the layers of  $\text{SiO}_2$ , the thin wafers of nanoporous silicon and the magnesium foils causes loss of material at rates of 23  $\text{nm day}^{-1}$ , 8  $\text{nm day}^{-1}$ , 9  $\mu\text{m day}^{-1}$  and 4  $\mu\text{m day}^{-1}$ , respectively, in artificial CSF (ACSF) at physiological temperature (37  $^{\circ}\text{C}$ ) (Supplementary Fig. 16). Separate studies indicate that PLGA (75:25 (lactide:glycolide) composition) dissolves in biofluids within four to five weeks<sup>19</sup>. To illustrate the various stages of dissolution of a completed system, Fig. 1k shows a sequence of images of a bioresorbable pressure sensor inserted into a transparent chamber designed for accelerated testing (polydimethylsiloxane (PDMS) enclosure filled with buffer solution at pH 12 and room temperature), in which fluid exchange can occur through an array of openings around the perimeter (Supplementary Fig. 17). Supplementary Fig. 18 presents images of nanoporous silicon and silicon nanomembranes observed by scanning electron microscopy at different times during hydrolysis. The silicon nanomembrane dissolves uniformly, without fracture. By comparison, nanoporous silicon dissolves less uniformly, with a tendency to form fragments. Here, the silicon-nanomembrane and  $\text{SiO}_2$  components dissolve first, within 15 hours, followed by the nanoporous silicon, which disappears within 30 hours. In all cases, the dissolution kinetics depends strongly on the materials and the composition of the surrounding solution<sup>20–22</sup>. Supplementary Table 1 summarizes the hydrolysis mechanisms and dissolution rates of these materials in a representative solution. As described below, the encapsulation material and its thickness define the operational lifetimes.

Figure 2 illustrates a strategy for using these types of bioresorbable systems for wireless pressure and temperature monitoring in the intracranial space of rats. Figure 2a shows a photograph of a device

like the one in Fig. 1c, but configured to allow simultaneous sensing of both pressure and temperature. The measured temperature can also be used to calibrate against parasitic effects of this parameter on the pressure determination (see Supplementary Methods and Supplementary Fig. 19). Biodegradable molybdenum wires (10  $\mu\text{m}$  thick) serve as an interface to wireless communication systems. Pressing the interconnect wires (molybdenum, 10  $\mu\text{m}$  thick, or magnesium, 50  $\mu\text{m}$  thick) against the PLGA at elevated temperatures (65  $^{\circ}\text{C}$ ) embeds them near the surface but leaves the top regions exposed, thereby allowing for deposition of biodegradable metals (molybdenum, 2  $\mu\text{m}$  thick) to form electrical contact pads through stencil masks (made from the polyimide Kapton, 12.5  $\mu\text{m}$  thick; Supplementary Fig. 20). The deposited molybdenum forms stable interconnects between metal wires and silicon nanomembranes that are fully embedded on PLGA. Encapsulation with a bioresorbable polymer (polyanhydride, discussed in more detail below) enhances system robustness by reducing the stress concentrations at the interconnections. Narrow strips of PLGA laminated onto the front and back sides of the wires along their entire lengths act as electrical insulation. These insulated wires connect to an externally mounted, miniaturized wireless potentiostat for transmission of data thorough percutaneous wiring. Figure 2b provides a diagram of such a system in the intracranial space of a rat model. The sensor subsystem connects via molybdenum wires to the wireless module, which is mounted on the top of the skull. Figure 2c–e summarizes the surgical process. A PLGA sheet (about 80  $\mu\text{m}$  thick) and a dissolvable surgical glue (Fig. 2c) seal the craniectomy defect to close the intracranial cavity. Conventional sutures hold the surgical site closed, in a standard process<sup>23</sup> that retains points at which the dissolvable wires emerge from the skin to allow electrical connection (Fig. 2d). These wires have dimensions comparable to those of the surgical threads, and therefore pose little additional risk. Figure 2e shows a healthy, freely moving rat with a complete system. Supplementary Fig. 21 presents images of the connections.

Figure 3 summarizes the results of a comprehensive set of wireless measurements of intracranial pressure (ICP) and intracranial



**Figure 3 | Wireless measurement of intracranial pressure and temperature with bioresorbable sensors implanted in live, freely moving animals.** **a–e**, Red, data from a transient, bioresorbable sensor; blue, data from a commercial sensor. **a**, Real-time wireless measurements of ICP, showing transient increases induced by the Valsalva manoeuvre. **b**, *In vivo* observation of changes in ICP as a function of time in the Trendelenburg and reverse Trendelenburg positions. ICP increases in the 30° head-down position (Trendelenburg) as compared with the supine position, and decreases in the 30° head-up position (reverse Trendelenburg). **c**, Gradual increase and **d**, decrease in ICT due to

temperature (ICT), recorded in rats with percutaneous wired systems. The ICP traces reveal features that correspond to periodic manual abdominal compression activating the Valsalva manoeuvre, which yields rapid increases or decreases in ICP (Fig. 3a)<sup>24</sup>. Gentle changes in the rat's position—that is, Trendelenburg (30° head-down position) and reverse Trendelenburg (30° head-up position)—produce gradual increases and decreases in ICP, respectively (Fig. 3b), as would be expected because of the corresponding accumulation and depletion of blood in the brain<sup>25</sup>. The pressure values compare well with those determined using a clinical-standard, wired ICP sensor implanted in the same region of the same animal. The wireless, bioresorbable ICT sensors perform to levels of accuracy similar to those of commercial sensors: Fig. 3c and d show comparative data collected by modulating the cranial temperature with a heating or cooling blanket placed beneath the animal.

The operational lifetimes of the devices are defined by dissolution of the encapsulation layers and the permeation of fluids through them. *In vitro* experiments using a bioresorbable pressure sensor

application of a heating/cooling blanket. **e**, Measurements of ICP over three days reveal consistent responses from devices encapsulated with biodegradable polyanhydride. **f**, Confocal fluorescence images of the cortical surface beneath the dissolved device at 2, 4 and 8 weeks, showing the absence of inflammatory responses. The images are double-immunostained for GFAP (glial fibrillary acidic protein) to detect astrocytes (red), and Iba1 (ionized calcium-binding adaptor molecule 1) to identify microglia/macrophages (green). The dashed line indicates the site of the implant.

encapsulated with a film of a specially synthesized polyanhydride (about 120 μm thick; Supplementary Fig. 22) show expected performance and accurate readings with an appropriately modified calibration factor ( $50 \Omega(\text{mm Hg})^{-1}$ ). The slow dissolution rate of the polyanhydride (about  $1.3 \mu\text{m day}^{-1}$ )—together with the modest change in sensitivity that occurs depending on the thickness of this material (about  $0.34 \Omega(\text{mm Hg})^{-1} \mu\text{m}^{-1}$ )—leads to a loss of accuracy of only a few per cent when operated over several days. This error falls within standards defined by the Association for the Advancement of Medical Instrumentation (AAMI) for pressure monitoring, that is,  $\pm 2 \text{ mm Hg}$  (from 0 to 20 mm Hg) and  $\pm 10\%$  (from 20 to 100 mm Hg)<sup>2</sup>. (Supplementary Methods and Supplementary Figs 23–25 present information on the synthesis/hydrolysis chemistry, dissolution kinetics, water permeability, and biocompatibility of the polyanhydride.)

Stable, continuous operation is possible for up to three days (Supplementary Fig. 26). Beyond this period, water tends to pass through the polyanhydride and PLGA into the electrically active regions of the device and the air cavity. The resistance remains

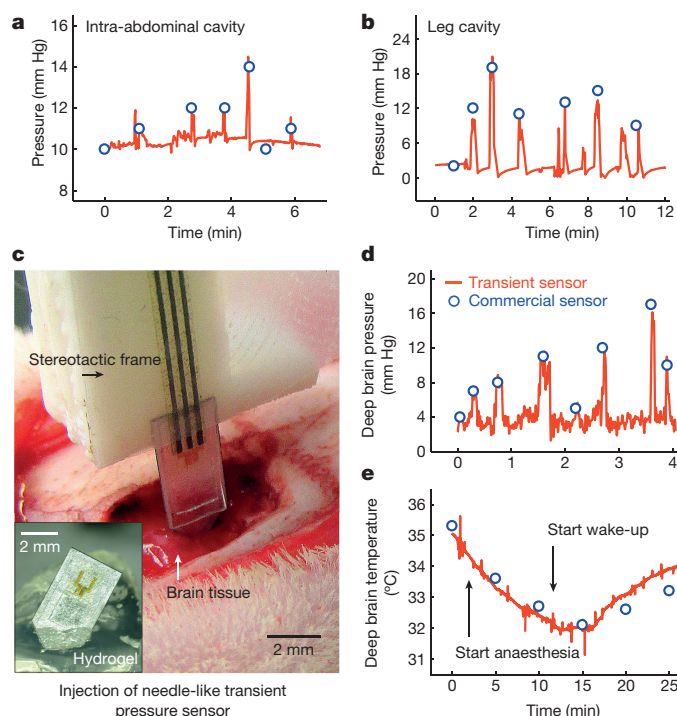


relatively constant for seven days, and then begins to increase markedly, mainly because of dissolution of the molybdenum wires and interconnection metal (Supplementary Figs 27 and 28). Figure 3e illustrates *in vivo* operation for three days without notable degradation in absolute accuracy or sensitivity, as benchmarked against a standard, non-resorbable wired sensor. Supplementary Fig. 29 shows similar data from the temperature sensor, where the absence of an air cavity affords enhanced stability, and accurate measurements for six days of operation. These timeframes are relevant for clinical use: ICP and ICT are typically monitored continuously for several days after traumatic brain injury. The chemistry, thickness and composition of the encapsulating layers can be selected to extend the functional lifetimes<sup>26</sup>.

Biocompatibility of the devices through all stages of their life cycle is essential. Comprehensive studies of the immunohistochemistry of brain tissues at several times after implantation (two, four and eight weeks) demonstrate that the sensors and the by-products of their dissolution in the intracranial space are biocompatible. Representative confocal fluorescence images (see Fig. 3f for nanoporous silicon and Supplementary Fig. 30 for magnesium foil) indicate no overt reaction of brain glial cells to the sensor, and no focal aggregation of glial cells at the implantation site for all time ranges. Astrocytosis (an increase in the number of astrocyte cells) and microglial activity at the cortical surface are within normal limits, indicating no overt immune reaction to the device and its by-products. Although the percutaneous wiring does not noticeably affect animal behaviour (see Supplementary Methods and Supplementary Fig. 31), a miniaturized, fully implantable wireless communication system might offer advantages, by removing the possibility of secondary infection at the wires. A wireless system constructed mostly, but not entirely, of resorbable materials (~85% by mass and ~86% by volume)—using an advanced near-field communication-technology approach, with fully bioresorbable metal coils, substrates and encapsulation layers—appears in Extended Data Fig. 1, Methods, and Supplementary Figs 32–37.

Given that these devices function successfully in the intracranial space, they could also be used in other organs and body compartments. As an example, Fig. 4a and b illustrate ICP monitoring using the same bioresorbable device in modes with relevance to acute abdominal compartment syndrome and acute compartment syndrome of the extremity<sup>27,28</sup>. Furthermore, modifying the devices to allow them to be injected deep into tissues could address other needs in clinical treatment. For example, monitoring physiological parameters of the deep brain with intraparenchymal sensors could yield data that are unavailable from the surface or the intracranial space. In addition, because electrophysiological and metabolic abnormalities often emanate from infarcts, contusions and haematomas that damage adjacent intact tissue, sensors of pressure, temperature, pH and other physical/chemical parameters that are placed into the parenchyma within the blood-deprived (ischaemic) penumbra could advance our knowledge of secondary brain injury<sup>29–31</sup>. Such considerations apply not only to injured brain tissue, but also to acute or chronic ischaemia that threatens the heart, limbs, intra-abdominal organs or grafts.

Modifying the geometry of the supporting structures introduced in Fig. 1 enables delivery of bioresorbable sensors into the depths of brain tissue, for direct measurements of injury or status. Figure 4c shows an example that integrates a bioresorbable ICP sensor onto a magnesium foil, formed with a tip region that allows injection into tissues of interest (Supplementary Fig. 38). Mounting the device on a stereotactic frame and fixture allows accurate positioning and controlled penetration (Supplementary Fig. 39). Figure 4d and e summarize pressure and temperature data collected at a site about 5 mm beneath the surface of the rat brain. The Valsalva manoeuvre yields data that quantitatively agree with those obtained using conventional sensors at a similar location. The device detected changes in temperature during anaesthesia (the temperature decreased, owing to reduced blood circulation) and waking up (the temperature returned to normal), as expected of intraparenchymal tissue.



**Figure 4 | Application of bioresorbable sensors to various body cavities, and demonstration of an injectable format for deep brain monitoring.** Red, data from a transient biodegradable sensor; blue, data from a commercial sensor. **a, b**, Pressures measured in **a**, intra-abdominal and **b**, leg cavities. **c**, Image showing *in vivo* injection of a needle-shaped biodegradable pressure sensor (using a magnesium foil support, ~80  $\mu$ m thick) into the brain parenchyma with a stereotactic frame and arm. The inset shows a biodegradable pressure sensor inserted into hydrogel, as evidence of the sensor's robust mechanical construction. **d**, *In vivo* measurements of pressure in the deep brain. **e**, *In vivo* measurements of temperature in the deep brain during anaesthesia. The temperature drops during anaesthesia owing to reduced blood circulation, and returns to normal after awakening.

The biomedical sensors reported here enable wireless data collection in body cavities and in deep tissues, with platforms that are fully bioresorbable, thereby allowing patients to be monitored until homeostasis has been achieved, and avoiding the risks associated with chronically implanted devices or their removal<sup>32,33</sup>. *In vivo* and *in vitro* experiments demonstrate precision measurements of pressure, temperature, motion, flow, thermal properties and pH, with possible extensions to biomolecular binding events. These features will be useful in diagnosing and treating a diverse range of medical conditions, from acute traumatic injuries such as extremity compartment syndrome, to chronic medical diseases such as diabetes. The materials, manufacturing methods and design layouts should be relevant to many other sensor modalities, with the potential for co-integration of advanced silicon-based integrated circuits, radio communication technologies, power supply and energy harvesters—each adapted from advances in transient electronics. Thus, it is realistic to expect that these devices could be used in sensing, recording, stimulating, and electrical control for medical monitoring and treatment, not only for the body regions explored here but also for areas such as the cardiac space and spinal system. Translating these technologies into clinical settings should provide patients and medical professionals with a vital set of tools for combating human disease.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 March; accepted 18 November 2015.

Published online 18 January 2016.



1. Poole, J. E. Present guidelines for device implantation: clinical considerations and clinical challenges from pacing, implantable cardiac defibrillator, and cardiac resynchronization therapy. *Circulation* **129**, 383–394 (2014).
2. Brain, T. F. Guidelines for the management of severe traumatic brain injury. VI. Indications for intracranial pressure monitoring. *J. Neurotrauma* **24**, S37–S44 (2007).
3. Chamis, A. L. *et al.* Staphylococcus aureus bacteremia in patients with permanent pacemakers or implantable cardioverter-defibrillators. *Circulation* **104**, 1029–1033 (2001).
4. Hall-Stoodley, L., Costerton, J. W. & Stoodley, P. Bacterial biofilms: from the natural environment to infectious diseases. *Nature Rev. Microbiol.* **2**, 95–108 (2004).
5. Maytin, M. & Epstein, L. M. Lead extraction is preferred for lead revisions and system upgrades. *Circ. Arrhythm. Electrophysiol.* **3**, 413–424 (2010).
6. Boutry, C. M. *et al.* Towards biodegradable wireless implants. *Phil. Trans. R. Soc. A* **370**, 2418–2432 (2012).
7. Ott, K. *et al.* Retained intracranial metallic foreign bodies: report of two cases. *J. Neurosurg.* **44**, 80–83 (1976).
8. Vajramani, G. V. *et al.* Persistent and intractable ventriculitis due to retained ventricular catheters. *Br. J. Neurosurg.* **19**, 496–501 (2005).
9. Hwang, S.-W. *et al.* A physically transient form of silicon electronics. *Science* **337**, 1640–1644 (2012).
10. Irimia-Vladu, M. “Green” electronics: biodegradable and biocompatible materials and devices for sustainable future. *Chem. Soc. Rev.* **43**, 588–610 (2014).
11. Bettinger, C. J. & Bao, Z. Organic thin film transistors fabricated on resorbable biomaterial substrates. *Adv. Mater.* **22**, 651–655 (2010).
12. Luo, M., Martinez, A. W., Song, C., Herrault, F. & Allen, M. G. A microfabricated wireless RF pressure sensor made completely of biodegradable materials. *J. Microelectromech. Syst.* **23**, 4–13 (2014).
13. Brogan, M. E. & Manno, E. M. Treatment of malignant brain edema and increased intracranial pressure after stroke. *Curr. Treat. Options Neurol.* **17**, 327 (2015).
14. Liu, C. *Foundations of MEMS* Ch. 6 (Prentice Hall, 2011).
15. Moseley, P. T. & Crocker, J. *Sensor Materials* Ch. 4 (Institute of Physics Publishing, 1994).
16. Chang, H. *et al.* DNA-mediated fluctuations in ionic current through silicon oxide nanopore channels. *Nano Lett.* **4**, 1551–1556 (2004).
17. Zheng, G., Patolsky, F., Cui, Y., Wang, W. U. & Lieber, C. M. Multiplexed electrical detection of cancer markers with nanowire sensor arrays. *Nature Biotechnol.* **23**, 1294–1301 (2005).
18. Stern, E. *et al.* Label-free immunodetection with CMOS-compatible semiconducting nanowires. *Nature* **445**, 519–522 (2007).
19. Gentile, P., Chiono, V., Carmagnola, I. & Hatton, P. V. An overview of poly(lactic-co-glycolic) acid (PLGA)-based biomaterials for bone tissue engineering. *Int. J. Mol. Sci.* **15**, 3640–3659 (2014).
20. Hwang, S.-W. *et al.* Dissolution chemistry and biocompatibility of single-crystalline silicon nanomembranes and associated materials for transient electronics. *ACS Nano* **8**, 5843–5851 (2014).
21. Kang, S.-K. *et al.* Dissolution behaviors and applications of silicon oxides and nitrides in transient electronics. *Adv. Funct. Mater.* **24**, 4427–4434 (2014).
22. Yin, L. *et al.* Dissolvable metals for transient electronics. *Adv. Funct. Mater.* **24**, 645–658 (2014).
23. Uslaner, J. M. *et al.* T-type calcium channel antagonism produces antipsychotic-like effects and reduces stimulant-induced glutamate release in the nucleus accumbens of rats. *Neuropharmacol.* **62**, 1413–1421 (2012).
24. Barth, K. N. M., Onesti, S. T., Krauss, W. E. & Solomon, R. A. A simple and reliable technique to monitor intracranial pressure in the rat: technical note. *Neurosurgery* **30**, 138–140 (1992).
25. Haure, P., Cold, G. E., Hansen, T. M. & Larsen, J. R. The ICP-lowering effect of 10° reverse Trendelenburg position during craniotomy is stable during a 10-minute period. *J. Neurosurg. Anesthesiol.* **15**, 297–301 (2003).
26. Morgan, P. W. Structure and moisture permeability of film-forming polymers. *Ind. Eng. Chem.* **45**, 2296–2306 (1953).
27. Meldrum, D. R. *et al.* Prospective characterization and selective management of the abdominal compartment syndrome. *Am. J. Surg.* **174**, 667–673 (1997).
28. Olson, S. A. & Glasgow, R. R. Acute compartment syndrome in lower extremity musculoskeletal trauma. *J. Am. Acad. Orthop. Surg.* **13**, 436–444 (2005).
29. Stiefel, M. F. *et al.* Reduced mortality rate in patients with severe traumatic brain injury treated with brain tissue oxygen monitoring. *J. Neurosurg.* **103**, 805–811 (2005).
30. Timofeev, I. *et al.* Extracellular brain pH with or without hypoxia is a marker of profound metabolic derangement and increased mortality after traumatic brain injury. *J. Cereb. Blood Flow Metab.* **33**, 422–427 (2013).
31. Suehiro, E. *et al.* Diverse effects of hypothermia therapy in patients with severe traumatic brain injury based on the computed tomography classification of the traumatic coma data bank. *J. Neurotrauma* **32**, 353–358 (2015).
32. Mittal, R. *et al.* Use of bio-resorbable implants for stabilization of distal radius fractures: the United Kingdom patients’ perspective. *Injury* **36**, 333–338 (2005).
33. Ye, T. *et al.* Management of grade III open dislocated ankle fractures: combined internal fixation with bioabsorbable screws/rods and external fixation. *J. Am. Podiatr. Med. Assoc.* **101**, 307–315 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** S.-K.K. and co-workers are funded by the Defense Advanced Research Projects Agency. J.G.M. is supported by the National Institute of Mental Health, grant F31MH101956. The authors thank M. R. Bruchas at Washington University School of Medicine for providing immunohistochemistry facilities; M. R. MacEwan at Washington University School of Medicine for discussions on animal protocols; A. Manocchi at Transient Electronics Inc. for performing the dissolution test of polyanhydride; and H. Ning at Xerion Advanced Battery Corporation for assistance with running the BET measurements. H.C. was a Howard Hughes Medical Institute International Student Research Fellow. S.-W.H. was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant NRF-2015R1C1A1A02037560). G.P. and K.M.L. were supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning (grants NRF-2007-00107 and NRF-2013M3A9D3045719).

**Author Contributions** S.-K.K., S.-W.H., D.V.H., N.A.K., S.Y., J.S., H.Y., R.C.W., C.H.L., S.C., D.S.W., J.C., P.V.B. and J.A.R. designed and fabricated the sensors and interfaces. S.-K.K., S.M.L., J.S., J.K. S.H.L. and J.A.R. designed, fabricated and analysed the near-field communication system with the sensor. S.-K.K., R.K.J.M., S.M.L., D.V.H., H.C., P.G., S.Y., J.S., M.S., R.C.W., C.H.L., B.V., Z.L., Y.H., W.Z.R. and J.A.R. conceived the idea and performed the experiments and analysis. R.K.J.M., P.G., J.G.M., M.S., G.P., A.D.G., A.H.K., K.-M.L. and W.Z.R. analysed the immunohistochemistry. S.-K.K., R.K.J.M., S.-W.H., S.M.L., D.V.H., H.C., W.Z.R. and J.A.R. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.R. ([jrogers@illinois.edu](mailto:jrogers@illinois.edu)) or W.Z.R. ([rayz@wudosis.wustl.edu](mailto:rayz@wudosis.wustl.edu)).

## METHODS

**Fabrication of bioresorbable silicon pressure sensors.** Fabrication involved integration of silicon-based, piezoresistive sensing elements onto substrates of PLGA, bonded over cavities etched into the surfaces of nanoporous Si (np-Si) substrates or magnesium foils. Solid-state diffusion of boron yielded highly doped p-type monocrystalline silicon nanomembranes (Si-NMs) on silicon-on-insulator (SOI) wafers (top silicon  $\sim 300$  nm thick, p-type; SOITEC, France). Eliminating the buried oxide with hydrofluoric acid allowed transfer of the Si-NMs onto a bilayer of D-PI (diluted polyimide (poly(pyromellitic dianhydride-co-4,4'-oxydianiline)),  $\sim 200$  nm)/PMMA (poly(methyl methacrylate),  $\sim 300$  nm) on temporary silicon carrier substrates. Photolithography and etching patterned the Si-NMs into structures with serpentine designs. Electron-beam evaporation and spin-casting defined uniform layers of SiO<sub>2</sub> ( $\sim 100$  nm) and D-PI, respectively, to serve the purpose of passivation. Selective dry etching through all of the layers (D-PI/SiO<sub>2</sub>/D-PI/PMMA) formed a mesh structure that enabled release in acetone, for transfer to a film of PLGA ( $\sim 30$   $\mu$ m). Heating these films to temperatures near the glass transition of the PLGA (65 °C) and laminating them onto np-Si substrates (or magnesium foils,  $\sim 60$ – $80$   $\mu$ m) with square regions of etched relief ( $\sim 30$ – $40$   $\mu$ m) formed sealed air cavities upon cooling to room temperature. Additional details appear in Supplementary Information.

**Calibration of the pressure response.** Responses of commercial sensors under environments similar to those in the intracranial cavity allowed absolute pressure calibration for the bioresorbable devices. The experiments involved placing a bioresorbable pressure sensor inside the barrel of a syringe partially filled with ACSF (Ecocyte BioScience, USA) and with a commercial sensor (NeuLog, USA) located at its open end (orifice). Moving the plunger component of the syringe allowed reversible access to well controlled pressures throughout a range relevant to intracranial monitoring (Supplementary Fig. 9). Comparison of the electrical resistance of the bioresorbable sensor (via data acquisition (DAQ) system USB-4065, National Instruments, USA) with pressures from the commercial sensor yielded calibration curves. Additional data appear in Supplementary Information.

**Connections to wireless data-transmission systems.** Laser cutting of foils of molybdenum ( $\sim 10$   $\mu$ m thick) or magnesium ( $\sim 50$   $\mu$ m thick) yielded dissolvable narrow metal strips (that is, interconnection wires,  $80$   $\mu$ m  $\times$   $30$  mm). Pressing these wires against PLGA substrates using a PDMS stamp at 65 °C embedded them into the surface of the PLGA. Sputter deposition of molybdenum ( $\sim 2$   $\mu$ m) through high-resolution stencil masks (12.5  $\mu$ m, Kapton; Dupont, USA) yielded electrical connections between the wires and contact pads on the PLGA (Supplementary Fig. 20). The opposite ends of the wires connected to externally mounted wireless communication systems (Pinnacle Technology, USA) (Supplementary Fig. 21).

**Evaluation of the kinetics of device dissolution.** Measurements of time-dependent changes in the thicknesses of square ( $100$   $\mu$ m  $\times$   $100$   $\mu$ m) Si-NMs ( $\sim 200$  nm thick), electron-beam evaporated layers of SiO<sub>2</sub> ( $\sim 100$  nm), free-standing nanoporous silicon substrates (np-Si,  $\sim 80$   $\mu$ m) and magnesium foils ( $\sim 80$   $\mu$ m) due to immersion in ACSF at body temperature (37 °C) established the dissolution kinetics of the key materials. Removing samples from the ACSF every other day, rinsing them with deionized water, and measuring the thicknesses by profilometry (Dektak, USA) yielded the dissolution rate, as in Supplementary Fig. 16. Sealed reservoirs of PDMS with viewing windows allowed for observation of dissolution behaviour at the level of the completed devices. These engineered structures included access channels around the periphery to allow passive fluid exchange and diffusion with a surrounding bath (Supplementary Fig. 17).

**Evaluation in animal models.** Studies were performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Institutional Animal Care and Use Committee (IACUC) of Washington University in St Louis (protocol number 20140207). Male Lewis rats weighing 250–350 g (Charles River, Wilmington, MA) received subcutaneous injections of buprenorphine hydrochloride ( $0.05$  mg kg<sup>-1</sup>; Reckitt Benckiser Healthcare Ltd, USA) for pain management, and of ampicillin ( $50$  mg kg<sup>-1</sup>; Sage Pharmaceuticals, USA) to prevent infection at the implantation site before the surgical process. Animals were anaesthetized with isoflurane gas and held in a stereotaxic frame for the duration of the surgical procedure and measurements. Opening a craniectomy and dural, implanting bioresorbable sensors on the cortical surface, sealing the craniectomy with a PLGA sheet ( $\sim 80$   $\mu$ m thick) and/or biodegradable surgical glue, and suturing the skin implanted the fully resorbable biosensing system in intracranial space. Comparison testing with a clinical intracranial pressure sensor (Integra LifeSciences, USA) and commercial thermistor (DigiKey Electronics, USA) implanted in parallel to bioresorbable sensors demonstrated the functionality of the bioresorbable sensors. To implant the injectable device, the same procedure of opening a craniectomy and dural was performed. Injecting needle-shaped biosensors into the brain parenchyma ( $\sim 5$  mm deep) with a stereotactic frame and arm enabled monitoring of

pressure and temperature in the deep-brain parenchyma. Additional details on the manual operation of pressure/temperature changes, the immunohistochemistry tests, and the surgical process and measurement at the intra-abdominal cavity and lower extremities appear in Supplementary Information. The immunohistochemistry tests used five individual rats per stage (2, 4 and 8 weeks) and device type (np-Si and magnesium-foil substrates). *In vivo* functionality tests of pressure and temperature sensors involved three trials using different batches of devices and animals, to establish reproducibility.

**Implantable near-field-communication wireless system.** The sensor introduced in Fig. 2a can be integrated with sub-dermal wireless data-transmission systems, constructed largely of bioresorbable materials, via thin, bioresorbable wires that pass through the skull. Extended Data Fig. 1a and b show an illustration of a chip-scale, near-field-communication (NFC) technology that includes bioresorbable coils, polymer substrates, encapsulation layers and resistors, a partially bioresorbable NFC chip, and non-resorbable capacitors, and a picture of this system integrated with a bioresorbable pressure sensor via biodegradable wiring. Here, micro-patterned magnesium coils ( $50$   $\mu$ m thick, outer diameter  $15$  mm) allow inductive coupling to an external data reader for power transfer and data transmission. A silicon-based logic chip (RF430FRL152H, Texas Instruments, USA;  $4$  mm  $\times$   $4$  mm  $\times$   $\sim 300$   $\mu$ m) captures the measured data at a high acquisition rate, then digitizes and processes the information for transmission to the external reader. Passive components include Si-NM resistors and capacitors. PLGA serves as the substrate and electrical passivation layer. Extended Data Fig. 1c summarizes the operating principles. The external reader wirelessly delivers power for operating the logic chip and provides the small currents needed to assess the response of the piezoresistive and thermoresistive sensors. In particular, changes in resistance associated with changes in pressure and temperature register as voltages that can be recorded and transmitted to the external reader by the NFC chip through the associated coil antenna. This NFC system is far more sophisticated than a conventional radio-frequency-identification (RFID) tag. Here, a single chip platform provides all of the computing functionality needed for high-speed data recording, real-time software filtering, and wireless transmission of sensor outputs as captured with an on-board 14-bit analogue-to-digital converter.

This system communicates through biofluids and tissue with little loss, owing to the use of magnetic coupling in a relatively low-frequency band (13.56 MHz; Supplementary Fig. 32), consistent with negligible heating associated with system operation (Supplementary Fig. 33). These characteristics enable communication distance of up to 25 mm through biological tissue. The high-speed, programmable operation of the NFC chip is critical to overall operation. Supplementary Fig. 34a presents examples of data-acquisition rates of up to 250 Hz, via recordings of oscillating voltages (sine wave) with frequencies from 1 Hz to 50 Hz. Spectrograms and other related data appear in Supplementary Fig. 34b–d. These high sampling rates allow efficient operation of digital filtering algorithms, and they also foreshadow the ability to measure biosignals such as EEG, ECoG and ECG. Supplementary Fig. 35 demonstrates the response of real-time high/low-pass filter function achieved by software programming for on-board computation with the NFC chip. Supplementary Fig. 36 shows two-channel operation of the system with/without this type of filtering during sensing. This integrated system provides wireless operation that compares quantitatively with that of a commercial wired sensor (Supplementary Fig. 37).

The wireless module is largely bioresorbable, as illustrated in Extended Data Fig. 1d through images at various stages of dissolution in ACSF at 60 °C. The magnesium coils, electrodes, interconnects and silicon resistors (240 mg;  $\sim 85\%$  of the total mass of the NFC system) dissolve fully after 14 days. Here, the NFC chip is not bioresorbable; but fully bioresorbable complementary metal-oxide semiconductor (CMOS) circuit technologies offer the potential for constructing bioresorbable chips. In particular, recently reported schemes demonstrate that modest modifications to otherwise conventional semiconductor-manufacturing techniques allow the use of foundry fabrication facilities for construction of bioresorbable CMOS<sup>34</sup>. Even with the examples presented here, where the NFC chip is not fully bioresorbable, the associated implantation strategy minimizes risk by locating the hardware subdermally on the skull, outside the intracranial space, thereby allowing rapid, facile extraction.

In this overall architecture, the fully bioresorbable sensors reside in the intracranial space, while the NFC system resides extracranially within the subgaleal layer of the scalp. Fine, dissolvable wiring provides electrical interconnections through a burr hole in the skull, sealed with a bioresorbable surgical glue. After completing the subgaleal closure, the wireless system and sensor are fully implanted. Extended Data Fig. 1e–h show a diagram of the implantation strategy, surgical process, and wireless *in vivo* intracranial pressure and temperature results measured in a rat model. Here, all of the components in the intracranial region are fully bioresorbable. The non-bioresorbable components of the system remain extra-axial within the

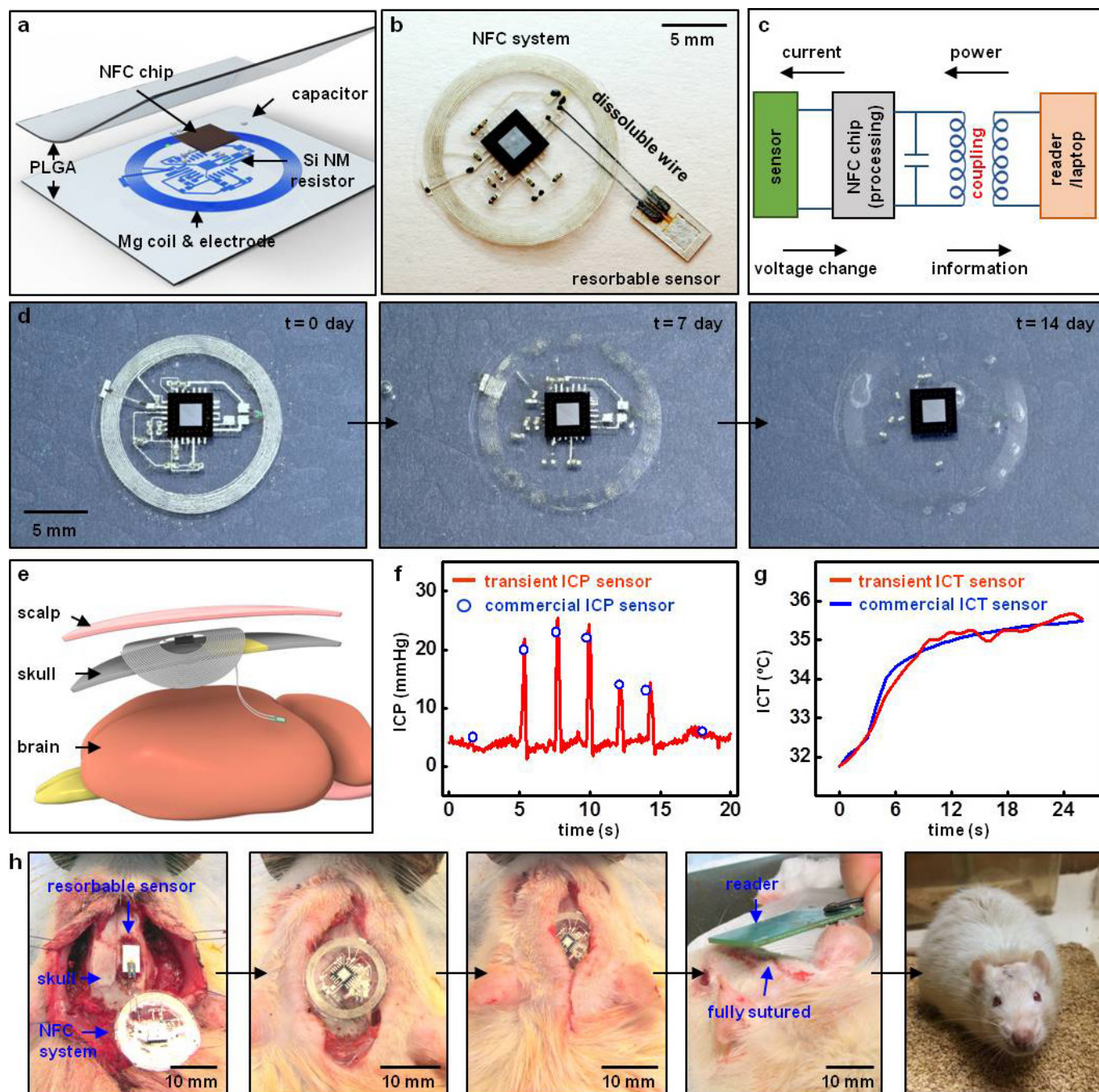
scalp, thereby minimizing the risk of provoking pathological neuroinflammation in the intracranial space. In addition, the relative material safety (as judged by the US Food and Drug Administration (FDA) class) of, for example, a subdermally implanted encapsulated non-resorbable device (such as an RFID chip) is similar to that of a titanium fixation screw. Removal of an extra-axial component involves a much lower risk than intracranial surgery. Intracranial pressure and temperature values measured in the rat model using the NFC system are comparable to those captured using commercial wired sensors.

**Fabrication of a fully implantable NFC wireless system.** The magnesium foil was patterned on the PDMS by using photolithography and etching with dilute hydrochloric acid (deionized water:HCl = 15:1). Transfer printing of the patterned magnesium foil onto a film of PLGA ( $\sim 150\mu\text{m}$ ) formed the inductive coil and

electrode. The Si-NM resistor was formed on the SOI wafer by doping with phosphorus at  $950^\circ\text{C}$  and patterning the top silicon ( $\sim 300\text{ nm}$  thick) into the trace. Undercutting the buried oxide with hydrofluoric acid and transfer printing Si-NM on PLGA formed the resistor of NFC system. Laminating the top PLGA ( $\sim 150\mu\text{m}$ ) and heating it at  $65^\circ\text{C}$  yielded the passivation layer. Biodegradable conductive W paste served to interconnect the NFC wireless system to the metal wire (molybdenum or magnesium)<sup>35</sup>.

34. Yin, L., Bozler, C., Harburg, D. V., Omenetto, F. & Rogers, J. A. Materials and fabrication sequences for water soluble silicon integrated circuits at the 90 nm node. *Appl. Phys. Lett.* **106**, 014105 (2015).
35. Huang, X. *et al.* Biodegradable materials for multilayer transient printed circuit boards. *Adv. Mater.* **26**, 7371–7377 (2014).





**Extended Data Figure 1 | Fully implantable near-field communication (NFC) system with bioresorbable interface and intracranial sensors.**

**a**, Diagram of a fully implantable NFC system. This device uses a magnesium foil ( $\sim 50\mu\text{m}$ ) for the inductive coil, interconnects and electrodes; patterned silicon nanomembranes (Si-NMs,  $\sim 300\text{ nm}$ ) for resistors; conventional capacitors; and an advanced NFC microchip for data acquisition, processing and transmission. PLGA serves as the substrate and for encapsulation. The diameter of the entire device is about 15 mm. **b**, Image of this type of NFC system integrated with a bioresorbable pressure sensor. **c**, Diagram of the operational principles. **d**, Series of images showing accelerated dissolution of the NFC system

inserted into an ACSF at  $60^\circ\text{C}$ . **e**, Diagram of the implantation process. The bioresorbable sensors reside in the intracranial space, while the NFC system is located extracranially, on the outside surface of the skull, beneath the skin. Bioresorbable, thin metal wires interconnect the NFC system and the sensors. **f**, Real-time wireless measurements of ICP, showing transient increases induced by the Valsalva manoeuvre (red, data obtained from a transient ICP sensor; blue, data obtained from a commercial ICP sensor). **g**, Increase in ICT owing to application of a heating blanket around the head, as determined by bioresorbable (red) and commercial (blue) sensors. **h**, Demonstrations of implantation and suturing in a rat model. A biodegradable surgical glue (TISSEAL) seals the intracranial space.



# Sub-particle reaction and photocurrent mapping to optimize catalyst-modified photoanodes

Justin B. Sambur<sup>1</sup>, Tai-Yen Chen<sup>1</sup>, Eric Choudhary<sup>1†</sup>, Guanqun Chen<sup>1</sup>, Erin J. Nissen<sup>2</sup>, Elayne M. Thomas<sup>3</sup>, Ningmu Zou<sup>1</sup> & Peng Chen<sup>1</sup>

The splitting of water photoelectrochemically into hydrogen and oxygen represents a promising technology for converting solar energy to fuel<sup>1,2</sup>. The main challenge is to ensure that photogenerated holes efficiently oxidize water, which generally requires modification of the photoanode with an oxygen evolution catalyst (OEC) to increase the photocurrent and reduce the onset potential<sup>3</sup>. However, because excess OEC material can hinder light absorption and decrease photoanode performance<sup>4</sup>, its deposition needs to be carefully controlled—yet it is unclear which semiconductor surface sites give optimal improvement if targeted for OEC deposition, and whether sites catalysing water oxidation also contribute to competing charge-carrier recombination with photogenerated electrons<sup>5</sup>. Surface heterogeneity<sup>6</sup> exacerbates these uncertainties, especially for nanostructured photoanodes benefiting from small charge-carrier transport distances<sup>1,7,8</sup>. Here we use super-resolution imaging<sup>9–13</sup>, operated in a charge-carrier-selective manner and with a spatiotemporal resolution of approximately 30 nanometres and 15 milliseconds, to map both the electron- and hole-driven photoelectrocatalytic activities on single titanium oxide nanorods. We then map, with sub-particle resolution (about 390 nanometres), the photocurrent associated with water oxidation, and find that the most active sites for water oxidation are also the most important sites for charge-carrier recombination. Site-selective deposition of an OEC, guided by the activity maps, improves the overall performance of a given nanorod—even though more improvement in photocurrent efficiency correlates with less reduction in onset potential (and vice versa) at the sub-particle level. Moreover, the optimal catalyst deposition sites for photocurrent enhancement are the lower-activity sites, and for onset potential reduction the optimal sites are the sites with more positive onset potential, contrary to what is obtainable under typical deposition conditions. These findings allow us to suggest an activity-based strategy for rationally engineering catalyst-improved photoelectrodes, which should be widely applicable because our measurements can be performed for many different semiconductor and catalyst materials.

Single-crystalline rutile titanium oxide (TiO<sub>2</sub>) nanorods (each of width  $127 \pm 27$  nm and length  $1,735 \pm 437$  nm) with defined surface facet orientations<sup>14</sup>, known to be stable photoanodes for water oxidation<sup>2</sup>, were dispersed on an indium-doped tin oxide (ITO) electrode in a microfluidic photoelectrochemical cell (Fig. 1a). To map the surface reactions of photogenerated holes and electrons on individual nanorods, we used single-molecule fluorescence microscopy<sup>9–12</sup> of either photogenerated hole ( $h^+$ )-induced amplex red oxidation or electron ( $e^-$ )-induced resazurin reduction (Extended Data Fig. 1) that both produce the fluorescent product resorufin as shown by ensemble photo (electro)catalysis measurements (Extended Data Figs 2, 3). In a typical experiment, the ITO electrode potential ( $E$ , referenced to the Ag/AgCl electrode) was fixed while a N<sub>2</sub>-purged pH 8.3 electrolyte solution

containing nanomolar nonfluorescent probe molecules flowed through the cell. Continuous 375-nm laser illumination generated charge-carriers within the TiO<sub>2</sub> nanorods, inducing a potential-dependent steady-state anodic photocurrent ( $i_{ss}$ , Fig. 1n, blue circles) from water oxidation<sup>2</sup>; the low probe concentration contributes insignificantly to  $i_{ss}$ . Another 532-nm laser induced the fluorescence of product molecules formed on nanorod surfaces. Anodic photocurrent and appreciable probe reactions were observed only with 375-nm laser illumination (Extended Data Figs 4a, 3c and 3e). Imaging and localizing individual resorufin molecules at a precision of about 30 nm (see Methods) maps the positions of hole- or electron-induced reactions across a range of potentials on each nanorod relative to its scanning electron microscope (SEM) image (Fig. 1b–g).

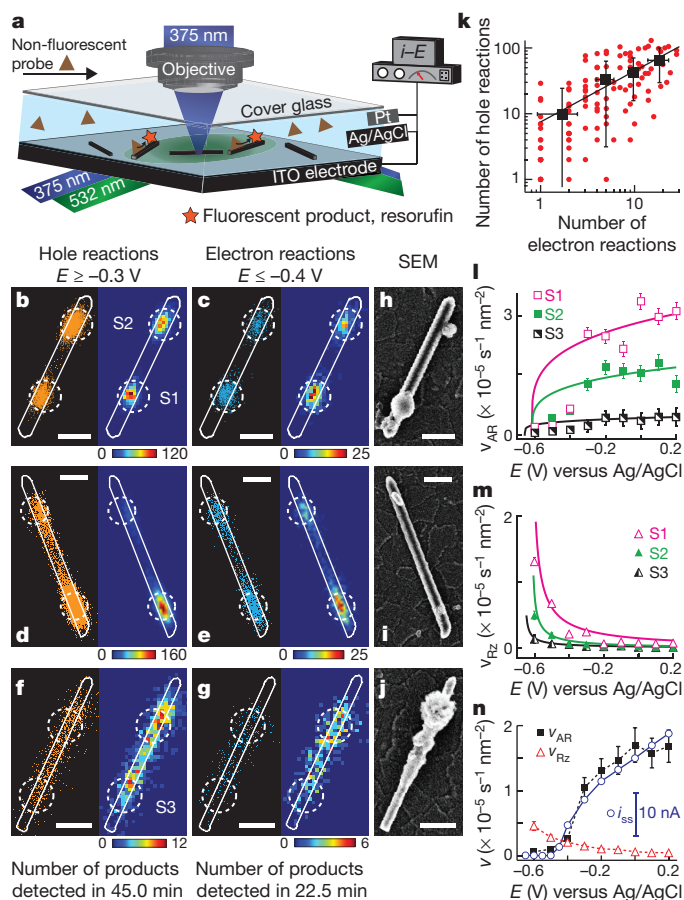
At potentials where  $i_{ss}$  is large ( $E \geq -0.3$  V, Fig. 1n), hole-induced surface reaction rates are high but strikingly non-uniform along single nanorods (Fig. 1b, d, f), even though identical {100} facets span the length of each nanorod<sup>14</sup>. For 25 out of 37 nanorods, hole-induced reactions predominantly occur in one or a few nanometre-scale ‘hot-spots’ (areas on a single nanorod denoting high hole activity) (Fig. 1b, d), while the remaining surface is less active. The smallest hotspots, of  $\sim 50$  nm (Extended Data Fig. 5b, c), are much smaller than the diffraction-limited resolution of  $\sim 300$  nm of our microscopic set-up, but are observable owing to super-resolution (that is, better than diffraction-limited optical resolution) as in earlier catalysis imaging studies<sup>9</sup>. For the remaining 12 out of the 37 nanorods studied, hole-induced reactions are distributed more evenly along the nanorods (Fig. 1f). At potentials where the anodic  $i_{ss}$  is minimal ( $E \leq -0.4$  V, Fig. 1n), the electron-induced reaction rates are high and also occur non-uniformly along the nanorods (Fig. 1c, e, g). We note that intra-facet heterogeneity in the photocatalytic activity of TiO<sub>2</sub> crystals has also been observed in earlier studies<sup>11,12</sup>.

The non-uniform distribution of reaction rates, and hence activity, indicates that the active site distribution differs among nanorods, and that the higher-activity sites are not {100} facet sites but instead probably the surface structural defects that we can discern using electron microscopy or the impurity atoms we detect in our elemental analysis (Extended Data Fig. 6). The reconstructed rutile {100} surface contains oxygen vacancies and small Ti<sub>2</sub>O<sub>3</sub> units, which could act as catalytic units for water oxidation<sup>15</sup>. Of the impurities present in our system, Fe has been identified as a dopant that increases both photocatalytic reduction and oxidation activity of TiO<sub>2</sub> nanoparticles<sup>16</sup>.

Surprisingly, our data indicate a strong spatial correlation between hole- and electron-induced activities (Fig. 1k): locations with high (or low) hole-induced activity at more positive potentials also exhibit high (or low) electron-induced activity at more negative potentials (see Fig. 1b versus Fig. 1c). The average separation between correlated hole- and electron-reaction hotspots on the same nanorod is  $\sim 40$  nm (Extended Data Fig. 5d), comparable to our localization precision. This spatial correlation is unlikely to arise from preferential probe

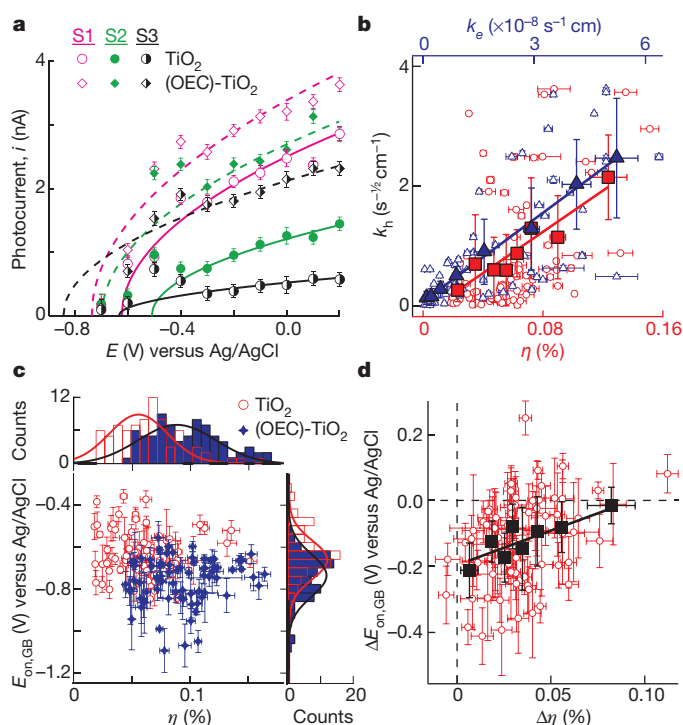
<sup>1</sup>Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, USA. <sup>2</sup>Department of Chemistry, Colorado Mesa University, Grand Junction, Colorado 81501, USA.

<sup>3</sup>Department of Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>†</sup>Present address: Center for Nanoscience and Technology, NIST, Gaithersburg, Maryland 20899, USA.



**Figure 1 | Super-resolution hole and electron surface reaction mapping.** **a**, Experimental setup for wide-field single-molecule fluorescence imaging of photoelectrocatalysis via two-laser total internal reflection excitation, or for sub-nanorod photocurrent measurements via focused laser excitation, in a three-electrode microfluidic photoelectrochemical cell (Methods). **b**, Scatter plot (left) and two-dimensional histogram ( $40 \times 40 \text{ nm}^2$  pixels; right) of all individual resorufin product molecules (orange dots) generated from hole-induced amplex red oxidation reactions at  $E \geq -0.3 \text{ V}$  on a single  $\text{TiO}_2$  nanorod with two localized hotspots. The solid white line is the outline of the nanorod, determined using SEM. Dashed white circles denote focused 375-nm laser spots for photocurrent measurements and OEC deposition. The colour scales indicate number of products formed. **c**, Same as **b**, but at  $E \leq -0.4 \text{ V}$ , and the blue dots are resorufin product molecules generated from electron-induced resazurin reduction reactions. **d–g**, Same as **b** and **c**, but for a nanorod with a single dominant hole reaction hotspot (**d**, **e**) or delocalized hole reactions (**f**, **g**). **h–j**, SEM images of the three nanorods in **b–g** after OEC deposition. All scale bars are 400 nm. **k**, Pixel-to-pixel correlation (red dots) between the two-dimensional histograms of hole and electron surface reactions, respectively, in **b** and **c** for the whole nanorod with a cross-correlation coefficient of  $\rho = 0.84 \pm 0.01$ . Black squares represent the binned and averaged data to obtain the general trend; the black line is the linear fit. **l** and **m**, Potential dependences of  $v_{\text{AR}}$  and  $v_{\text{Rz}}$  at spots S1, S2 and S3 on the nanorods in **b**, **c**, **f** and **g**. Solid lines are global fits with equation (2) (for  $E > -0.4 \text{ V}$ ) and equation (3). **n**, Comparison of  $i_{\text{ss}}$  from  $\sim 1,000$   $\text{TiO}_2$  nanorods and nanorod-averaged  $v_{\text{AR}}$  and  $v_{\text{Rz}}$  for 78 spots from 37 nanorods. All error bars represent s.d.

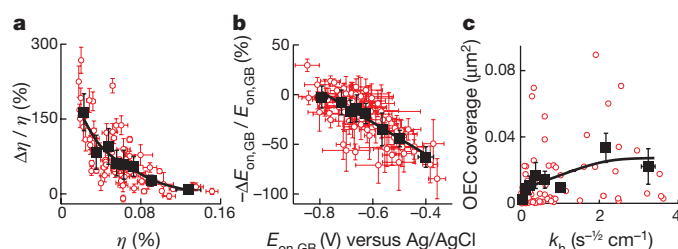
adsorption to the same surface sites, given that controls show the structurally similar product (Extended Data Fig. 1a) adsorbs equivalently to all surface sites (Extended Data Fig. 7). We note that although it has been shown that photogenerated holes and electrons can preferentially react at the same facet of semiconductor microcrystal photocatalysts<sup>17</sup>, our data further indicate that holes and electrons tend to reach and react at the same surface sites within a given nanorod facet.



**Figure 2 | Correlation between local  $i$ - $E$  responses, surface activities, and OEC effects.** **a**,  $i$ - $E$  curves from spots S1, S2 and S3 in Fig. 1b, c, f and g before and after OEC deposition. Lines are fits using equation (1). **b**, Correlation of  $k_{\text{h}}$  with  $\eta$  and with  $k_{\text{e}}$  for the same 78 spots (open symbols), with cross-correlation coefficients  $\rho(k_{\text{h}}, \eta) = 0.45 \pm 0.06$  and  $\rho(k_{\text{h}}, k_{\text{e}}) = 0.75 \pm 0.03$ . Solid symbols are binned and averaged data to obtain general trends; lines are linear fits. **c**,  $E_{\text{on,GB}}$  versus  $\eta$  for 78 nanorod spots before and after OEC deposition. Top and right, histograms of  $\eta$  and  $E_{\text{on,GB}}$ . Counts are the number of nanorod spots. Lines are Gaussian fits, giving:  $\eta = 0.056 \pm 0.023\%$  and  $E_{\text{on,GB}} = -0.66 \pm 0.09 \text{ V}$  for  $\text{TiO}_2$  and  $\eta_{\text{OEC}} = 0.088 \pm 0.033\%$  and  $E_{\text{on,GB,OEC}} = -0.73 \pm 0.10 \text{ V}$  for (OEC)- $\text{TiO}_2$ . **d**, Correlation between  $\Delta E_{\text{on,GB}} = E_{\text{on,GB,OEC}} - E_{\text{on,GB}}$  and  $\Delta \eta = \eta_{\text{OEC}} - \eta$  for 78 spots (red circles). Black squares are data binned and averaged to obtain the general trend; the black line is the linear fit. Error bars for individual nanorod spots represent s.d. and error bars for binned and averaged data represent s.e.m.

At any location on a nanorod, whether a hotspot (see S1 in Fig. 1b) or a low-activity spot (see S3 in Fig. 1f), the hole-induced amplex red oxidation rate ( $v_{\text{AR}}$ ) becomes clearly noticeable at about  $-0.5 \text{ V}$  (Fig. 1l), commensurate with the steady-state photocurrent  $i_{\text{ss}}$  and its onset potential ( $E_{\text{on,ss}}$ ) of  $-0.47 \text{ V}$  (Fig. 1n). We find that  $v_{\text{AR}}$  scales with  $E^{3/4}$  for  $E \geq -0.3 \text{ V}$  (Extended Data Fig. 3b) and with  $I_0^{1/2}$  when using a fixed positive potential ( $I_0 = 375 \text{ nm}$  laser power; Extended Data Fig. 3c), consistent with amplex red reacting with surface-adsorbed  $\text{OH}^{\bullet}$  radicals produced from hole-induced oxidation of  $\text{OH}^-$  or  $\text{H}_2\text{O}$  (ref. 18). The electron-induced resazurin reduction rate ( $v_{\text{Rz}}$ ) shows an opposite trend:  $v_{\text{Rz}}$  is high at  $E < E_{\text{on,ss}}$  (Fig. 1m) and decreases greatly at  $E > E_{\text{on,ss}}$ , where a  $\sim 0.02\text{-nA}$  ( $\sim 10^8$  holes per second) anodic photocurrent is measured per nanorod (Fig. 1n). The measurable  $v_{\text{Rz}}$  at  $E > E_{\text{on,ss}}$  indicates that even under highly anodic conditions, single-molecule fluorescence microscopy is able to detect the photogenerated electrons that escape being collected by the ITO electrode and participate instead in surface reactions (that is, surface recombination).

To assess local water oxidation efficiencies, we focused a 375-nm laser (focus diameter  $\sim 390 \text{ nm}$ ) onto a nanorod spot and measured potential-dependent photocurrents through the ITO electrode (Fig. 1a). Different  $i$ - $E$  responses measured for different spots on the same nanorod (such as S1 versus S2 in Fig. 2a) suggest that we are indeed measuring local, sub-nanorod photoelectrochemical behaviour. We find that the anodic photocurrent  $i$  always scales with  $(E)^{1/2}$



**Figure 3** |  $\eta$  and  $E_{\text{on,GB}}$  changes of OEC-modified  $\text{TiO}_2$  nanorods.

**a**, Relative change in absorbed-photon-to-current efficiency ( $\Delta\eta/\eta$ ) versus initial  $\eta$  for 78 nanorod spots (open circles). **b**, Relative change in onset potential ( $-\Delta E_{\text{on,GB}}/E_{\text{on,GB}}$ ) versus initial  $E_{\text{on,GB}}$  (open circles), where negative values represent negative shifts in  $E_{\text{on,GB}}$ . **c**, Catalyst coverage (Methods) versus  $k_h$ . Black squares show binned and averaged data; black lines are spline interpolations. Error bars for individual nanorod spots represent s.d. and error bars for binned and averaged data represent s.e.m.

for  $E \geq -0.3$  V (Extended Data Fig. 8d); for  $E < -0.3$  V and especially near  $E_{\text{on,ss}}$  ( $-0.47$  V), it deviates from the  $(E)^{1/2}$  dependence owing to photocurrent transient dynamics (Extended Data Fig. 4b)<sup>19</sup>. At fixed positive potentials, the photocurrent scales linearly with  $I_0$  (Extended Data Fig. 4c). Spots with higher hole and electron activities (such as spot S1) show larger photocurrents, indicating that higher surface activity is related to higher water oxidation efficiency.

To relate local water oxidation efficiency to surface hole and electron activities, we use the Gärtner–Butler<sup>20,21</sup> and Reichman<sup>22</sup> models to analyse quantitatively the  $i$ - $E$ ,  $v_{\text{AR}}$ - $E$ , and  $v_{\text{RZ}}$ - $E$  data from 78 nanorod spots (equations (1) to (3) in the Methods). For each spot, this yields  $\eta$  (the absorbed-photon-to-current efficiency within the depletion layer of the nanorod),  $E_{\text{on,GB}}$  (the photocurrent onset potential predicted by the Gärtner–Butler model) and the effective rate constants  $k_h$  and  $k_e$  that represent hole and electron surface activities, respectively. Spots with larger  $\eta$  have larger  $k_h$  (Fig. 2b), as expected from the association between water oxidation efficiency and surface hole activity, with both terms including the efficiencies of hole transport to the surface and interfacial charge transfer<sup>23</sup>. The  $k_h$  value for each spot is also strongly correlated with its  $k_e$  value, further supporting the idea that photogenerated holes and electrons prefer to reach and react at the same sites. Importantly, the strong correlation between  $\eta$  and surface hole and electron activities suggests that the nanorod sites exhibiting higher water oxidation efficiency effectively mediate both oxidation and reduction reactions (the dominant process depends on the applied potential), making them also more effective surface recombination sites than the lower efficiency sites. This dual oxidation–reduction activity is consistent with the earlier observation that both oxidative and reductive photocatalytic reactions preferentially occur on the same {100} facets on  $\text{SrTiO}_3$  crystals<sup>17</sup>. We hypothesize that this dual activity might make it easier for high-efficiency sites to accept photogenerated holes from the semiconductor interior and transfer them to adsorbed  $\text{OH}^-$  or  $\text{H}_2\text{O}$  (that is, the sites are oxidized and reduced sequentially in a redox cycle, illustrated in Supplementary Scheme 1).

In a next step, we photoelectrochemically deposited the OEC cobalt–borate ( $\text{Co-B}_i$ ) site-selectively by focusing a 375-nm laser onto either high- or low-activity spots where  $i$ - $E$  data had been measured. This resulted in heterogeneous OEC coverage (Fig. 1h–j) and enhanced photocurrents in subsequent  $i$ - $E$  measurements (Fig. 2a), as expected for this amorphous catalyst<sup>24,25</sup>, with the enhancement saturating<sup>26</sup> as increasing catalyst is deposited (Extended Data Fig. 9b). Gärtner–Butler model analysis shows that OEC deposition increases the value of  $\eta$  for nearly all nanorod spots (that is,  $\Delta\eta > 0$ , Fig. 2d), on average by  $\sim 60\%$  (Fig. 2c, top). OEC deposition also induced an average shift in  $E_{\text{on,GB}}$  of  $\sim 70$  mV (Fig. 2c, right), as expected from the reduced overpotential seen in OEC-modified photoanodes<sup>27</sup>, but surprisingly 16% of spots exhibited positive shifts (that is,  $\Delta E_{\text{on,GB}} > 0$ ; Fig. 2d). For spots exhibiting the expected  $\Delta E_{\text{on,GB}} < 0$  V, the magnitude of  $\Delta E_{\text{on,GB}}$  is smaller for

spots with larger  $\Delta\eta$  (Fig. 2d). We conclude that photocurrent enhancements from OEC deposition on these nanorods are not necessarily accompanied by negative shifts in  $E_{\text{on,GB}}$ , and that the magnitudes of their changes can be anticorrelated.

The most striking effects of OEC deposition are changes in  $\eta$  and  $E_{\text{on,GB}}$  compared with their initial values. A strong negative correlation exists for the relative change in  $\eta$  (that is,  $\Delta\eta/\eta$ , Fig. 3a): spots with smaller initial  $\eta$ , characterized also by lower surface hole and electron activities, experience larger relative enhancements that can exceed 200%. This trend could be partly due to larger gains in interfacial charge transfer efficiency for initially low- $\eta$  spots, because OEC does not typically affect charge transport<sup>26</sup>. A strong negative correlation also exists between  $\Delta E_{\text{on,GB}}/E_{\text{on,GB}}$  and  $E_{\text{on,GB}}$  (Fig. 3b): spots with more positive initial  $E_{\text{on,GB}}$  exhibit larger relative negative shifts in  $E_{\text{on,GB}}$ . We attribute this trend to the ability of the OEC to relieve excess positive surface charge<sup>26,28</sup> and that larger negative shifts in  $E_{\text{on,GB}}$  are possible for sites with initially more excess surface charge<sup>29</sup>.

OEC deposition sites that optimize photocurrent enhancement are thus those with the lowest hole and electron activities, while deposition sites that most reduce onset potentials are those with the most positive  $E_{\text{on,GB}}$  values. These two types of sites are not necessarily the same (Fig. 2d), which highlights the challenge of trying to engineer photoanodes with a minimal amount of catalyst. First, under typical photoelectrodeposition (or photodeposition) conditions, more catalyst material is deposited onto higher-activity sites (Fig. 3c and refs 30, 31) that show less photocurrent enhancement. Second, upon scanning the potential positively, OEC deposition is expected to initiate earlier at sites with more negative  $E_{\text{on,GB}}$ . Both effects deposit more catalyst where it is least needed.

We propose to circumvent this problem with a block–deposit–remove strategy based on function rather than structure (Extended Data Fig. 10). This would involve first blocking sites with more negative  $E_{\text{on,GB}}$  and higher  $\eta$  by photoelectrochemically depositing protecting groups at increasingly positive potentials; then depositing the desired catalyst at the remaining sites with more positive  $E_{\text{on,GB}}$  or lower  $\eta$ ; and, finally, removal of the blocking groups to yield a photoanode with optimally located catalysts.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 20 July; accepted 7 December 2015.**

- Walter, M. G. *et al.* Solar water splitting cells. *Chem. Rev.* **110**, 6446–6473 (2010).
- Fujishima, A. & Honda, K. Electrochemical photolysis of water at a semiconductor electrode. *Nature* **238**, 37–38 (1972).
- Yang, J., Wang, D., Han, H. & Li, C. Roles of cocatalysts in photocatalysis and photoelectrocatalysis. *Acc. Chem. Res.* **46**, 1900–1909 (2013).
- Carroll, G., Zhong, D. & Gamelin, D. Mechanistic insights into solar water oxidation by cobalt-phosphate-modified  $\alpha\text{-Fe}_2\text{O}_3$  photoanodes. *Energy Environ. Sci.* **8**, 577–584 (2015).
- Wilson, R. H. A model for the current-voltage curve of photoexcited semiconductor electrodes. *J. Appl. Phys.* **48**, 4292–4297 (1977).
- Buurmans, I. L. C. & Weckhuysen, B. M. Heterogeneities of individual catalyst particles in space and time as monitored by spectroscopy. *Nature Chem.* **4**, 873–886 (2012).
- Liu, C., Dasgupta, N. P. & Yang, P. Semiconductor nanowires for artificial photosynthesis. *Chem. Mater.* **26**, 415–422 (2014).
- Warren, S. C. *et al.* Identifying champion nanostructures for solar water-splitting. *Nature Mater.* **12**, 842–849 (2013).
- Roelfaers, M. B. J. *et al.* Super-resolution reactivity mapping of nanostructured catalyst particles. *Angew. Chem. Int. Ed.* **48**, 9285–9289 (2009).
- Zhou, X. *et al.* Quantitative super-resolution imaging uncovers reactivity patterns on single nanocatalysts. *Nature Nanotechnol.* **7**, 237–241 (2012).
- Zhang, Y. *et al.* Superresolution fluorescence mapping of single-nanoparticle catalysts reveals spatiotemporal variations in surface reactivity. *Proc. Natl Acad. Sci. USA* **112**, 8959–8964 (2015).
- Tachikawa, T., Yamashita, S. & Majima, T. Evidence for crystal-face-dependent  $\text{TiO}_2$  photocatalysis from single-molecule imaging and kinetic analysis. *J. Am. Chem. Soc.* **133**, 7197–7204 (2011).
- Xu, W. *et al.* Single-molecule electrocatalysis by single-walled carbon nanotubes. *Nano Lett.* **9**, 3968–3973 (2009).



14. Liu, B. *et al.* Large-scale synthesis of transition-metal-doped TiO<sub>2</sub> nanowires with controllable overpotential. *J. Am. Chem. Soc.* **135**, 9995–9998 (2013).
15. Salvador, P., Garcia Gonzalez, M. L. & Munoz, F. Catalytic role of lattice defects in the photoassisted oxidation of water at (001) n-TiO<sub>2</sub> rutile. *J. Phys. Chem.* **96**, 10349–10353 (1992).
16. Choi, W., Termin, A. & Hoffmann, M. R. The role of metal ion dopants in quantum-sized TiO<sub>2</sub>: correlation between photoreactivity and charge carrier recombination dynamics. *J. Phys. Chem.* **98**, 13669–13679 (1994).
17. Giocondi, J., Salvador, P. & Rohrer, G. The origin of photochemical anisotropy in SrTiO<sub>3</sub>. *Top. Catal.* **44**, 529–533 (2007).
18. Villarreal, T. L., Gómez, R., Neumann-Spallart, M., Alonso-Vante, N. & Salvador, P. Semiconductor photooxidation of pollutants dissolved in water: a kinetic model for distinguishing between direct and indirect interfacial hole transfer. I. Photoelectrochemical experiments with polycrystalline anatase electrodes under current doubling and absence of recombination. *J. Phys. Chem. B* **108**, 15172–15181 (2004).
19. Salvador, P. Kinetic approach to the photocurrent transients in water photoelectrolysis at n-TiO<sub>2</sub> electrodes. 1. Analysis of the ratio of the instantaneous to steady-state photocurrent. *J. Phys. Chem.* **89**, 3863–3869 (1985).
20. Gärtner, W. W. Depletion-layer photoeffects in semiconductors. *Phys. Rev.* **116**, 84–87 (1959).
21. Butler, M. A. Photoelectrolysis and physical properties of the semiconducting electrode WO<sub>3</sub>. *J. Appl. Phys.* **48**, 1914–1920 (1977).
22. Reichman, J. Collection efficiency of low-mobility solar cells. *Appl. Phys. Lett.* **38**, 251–253 (1981).
23. Dotan, H., Sivula, K., Gratzel, M., Rothschild, A. & Warren, S. C. Probing the photoelectrochemical properties of hematite ( $\alpha$ -Fe<sub>2</sub>O<sub>3</sub>) electrodes using hydrogen peroxide as a hole scavenger. *Energy Environ. Sci.* **4**, 958–964 (2011).
24. Surendranath, Y., Lutterman, D. A., Liu, Y. & Nocera, D. G. Nucleation, growth, and repair of a cobalt-based oxygen evolving catalyst. *J. Am. Chem. Soc.* **134**, 6326–6336 (2012).
25. Khnayzer, R. S. *et al.* Structure and activity of photochemically deposited “CoPi” oxygen evolving catalyst on titania. *ACS Catal.* **2**, 2150–2160 (2012).
26. Klahr, B., Gimenez, S., Fabregat-Santiago, F., Bisquert, J. & Hamann, T. W. Photoelectrochemical and impedance spectroscopic investigation of water oxidation with “Co–Pi”-coated hematite electrodes. *J. Am. Chem. Soc.* **134**, 16693–16700 (2012).
27. Gamelin, D. R. Water splitting: catalyst or spectator? *Nature Chem.* **4**, 965–967 (2012).
28. Lin, F. & Boettcher, S. W. Adaptive semiconductor/electrocatalyst junctions in water-splitting photoanodes. *Nature Mater.* **13**, 81–86 (2013).
29. Peter, L. M., Li, J. & Peat, R. Surface recombination at semiconductor electrodes: Part I. Transient and steady-state photocurrents. *J. Electroanal. Chem. Interfacial Electrochem.* **165**, 29–40 (1984).
30. Steinmiller, E. M. P. & Choi, K. S. Photochemical deposition of cobalt-based oxygen evolving catalyst on a semiconductor photoanode for solar oxygen production. *Proc. Natl Acad. Sci. USA* **106**, 20633–20636 (2009).
31. Li, R. *et al.* Spatial separation of photogenerated electrons and holes among {010} and {110} crystal facets of BiVO<sub>4</sub>. *Nature Commun.* **4**, 1432 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The research is supported by the Department of Energy, Office of Science, Basic Energy Science (grant number DE-FG02-10ER16199), and in part by Army Research Office (grant numbers 63767-CH and 65814-CH), National Science Foundation (grant numbers CBET-1263736 and CHE-1137217 (to J.B.S.)), and REU programme DMR-1063059 (to E.J.N. and E.M.T.)), and the Petroleum Research Foundation (grant number 54289-ND7). This work used the Cornell Center for Materials Research Shared Facilities, an NSF MRSEC programme (grant number DMR-1120296), and the Cornell NanoScale Facility, a member of the NSF National Nanotechnology Infrastructure Network (grant number ECCS-15420819). We thank D. Hill of the Cahoon Group at the University of North Carolina at Chapel Hill for the finite-difference frequency-domain simulations presented in Supplementary Fig. 24. We also thank J. McKone, S. Maldonado, B. A. Parkinson, H. D. Abruña and A. J. Nozik for discussions, J. Colson and W. Dichtel for X-ray diffraction measurements, J. Grazul for electron microscopy and H. D. Abruña for electrochemical instrumentation.

**Author Contributions** J.B.S. and P.C. designed research. J.B.S. performed research. T.-Y.C., E.C., G.C., E.J.N., E.M.T. and N.Z. contributed to research. J.B.S. and P.C. analysed data and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C. ([pc252@cornell.edu](mailto:pc252@cornell.edu)).

## METHODS

**Synthesis and characterization of TiO<sub>2</sub> nanorods.** Single-crystalline rutile TiO<sub>2</sub> nanorods were synthesized via a molten-flux salt method following Liu *et al.*<sup>14</sup> and characterized by electron microscopy and X-ray diffraction. They have rectangular cross-sections with {100} side facets and {011} end facets. The ensemble-average diameter and length are  $62 \pm 16$  nm and  $263 \pm 113$  nm, respectively (Supplementary Fig. 1). The 37 nanorods on which we performed measurements have diameters of  $127 \pm 27$  nm and lengths of  $1,735 \pm 437$  nm.

**Redox-selective fluorogenic probe reactions.** The fluorogenic probe reaction was either oxidation of amplex red, a hole-probe, or reduction of resazurin, an electron-probe, both generating the fluorescent product resorufin (Extended Data Fig. 1a). Amplex red's oxidation potential is  $\sim 2$  V more negative than the valence band edge of TiO<sub>2</sub>, and resazurin's reduction potential is  $\sim 0.5$  V more positive than the conduction band edge of TiO<sub>2</sub>, determined from cyclic voltammetry (Extended Data Fig. 1b, c).

**Single-molecule super-resolution imaging of surface reactions.** A schematic illustration of the single-molecule super-resolution fluorescence microscopy experimental setup is shown in Fig. 1a, based on a wide-field prism-type total internal reflection fluorescence (TIRF) microscope as described previously<sup>10</sup>. TiO<sub>2</sub> nanorods were spin-coated onto an ITO electrode and annealed at 450 °C for 30 min, then assembled into an electrochemical microfluidic cell ( $\sim 5$  mm wide and 100  $\mu$ m high) using double-sided tape sandwiched between an ITO electrode and a coverslip (Supplementary Fig. 3). A syringe pump drives the continuous supply of the reactant solution at  $25 \mu\text{l min}^{-1}$ , which supplies  $1 \times 10^{10}$  probe molecules (at 50 nM) per second to the cell, whereas the maximum rate of product formation (or reactant consumption) from all of the nanorods in the flow cell was only  $5 \times 10^7$  molecules per second, estimated from the amplex red oxidation reaction rate at +0.2 V and the fact that  $\sim 2\%$  of the  $1.27\text{-cm}^2$  ITO surface area was covered with nanorods. In all experiments, the electrolyte solution was deaerated 1 M KCl, 100 mM pH 8.3 sodium borate buffer. The deaeration via N<sub>2</sub>-purging here removes O<sub>2</sub> from the cell, preventing both the competition for the photogenerated electrons by O<sub>2</sub> and the subsequent formation of O<sub>2</sub><sup>•−</sup>, which could oxidize amplex red<sup>32</sup>. All experiments were performed at room temperature.

A continuous wave circularly polarized 532-nm laser ( $630 \text{ W cm}^{-2}$ ) induces the fluorescence of the reaction product resorufin. The TiO<sub>2</sub> nanorods were also excited via TIRF illumination using a 375-nm laser ( $60 \text{ W cm}^{-2}$ ). The fluorescence was imaged by an electron multiplying charge coupled device (EMCCD) camera operated at a 15-ms frame rate. In a typical experiment, a sequence of fluorescence images (that is, 'catalytic movie') was acquired while the electrode potential, light intensity, flow rate and reactant concentration (either 50 nM amplex red or 50 nM resazurin) were fixed. Potential-dependent movies were acquired by stepping the potential in 0.1 V increments from  $-0.6$  V to  $+0.2$  V for amplex red oxidation experiments and from  $+0.2$  V to  $-0.6$  V for resazurin reduction experiments. The sequence of reactions were alternated (that is, amplex red imaging before resazurin imaging or vice versa for five different sample areas) and did not affect the results. The total acquisition time at each potential was 30,000 frames (that is, 7.5 min). As we only studied isolated nanorods, charge transport between adjacent nanorods does not apply to the photoelectrochemical processes measured.

The fluorescence images in the movies were analysed using a home-written MATLAB program, as described<sup>33</sup>, to localize the positions of individual fluorescent products on single nanorods to  $\sim 30$  nm precision (Extended Data Fig. 5a), in correlation with the nanorod's SEM image (for example, Fig. 1b, left). These positions can then be histogrammed to generate an image at super-optical resolution (for example, Fig. 1b, right). Detailed data analysis procedures can be found in Supplementary Note 2.

There are other techniques also capable of measuring electrochemical or catalytic activities of single nano-entities (see review<sup>34</sup>), including scanning electrochemical (cell) microscopy<sup>35–39</sup>, directly attaching a single particle to an ultramicroelectrode<sup>38,40</sup>, electrocatalytic current amplification via single-nanoparticle collisions<sup>41</sup>, surface plasmon resonance spectroscopy<sup>42–44</sup>, and conventional optical microscopy<sup>6,45</sup>.

**Elemental analysis and high-resolution electron microscopy.** The non-uniform hole and electron reaction rates along the single nanorods indicates that the higher-activity sites are most probably not the {100} facet sites. From the elemental and transmission electron microscopy (TEM) analysis (Extended Data Fig. 6), we hypothesize that these higher-activity sites are probably impurity atoms or surface structural defects.

For impurities, we have detected a number of impurity atoms in these nanorods via inductively coupled plasma atomic emission spectroscopy (ICP-AES, Extended Data Fig. 6a), among which Fe has been identified as a transition-metal dopant that

increases both photocatalytic reduction and oxidation activity of TiO<sub>2</sub> nanoparticles<sup>16</sup>. Yang and co-workers<sup>14</sup>, whose procedure we followed to synthesize these TiO<sub>2</sub> nanorods, intentionally doped these nanorods with 2% Fe, which resulted in a reduction in the overpotential for oxygen evolution on these nanorods, directly supporting our hypothesis. Elemental mapping of single nanorods further showed that some of these impurity atoms, including Fe, are heterogeneously distributed within a single nanorod (Extended Data Fig. 6b–f). There is also evidence in the literature that single impurity atoms can act as highly active catalytic sites on TiO<sub>2</sub> nanomaterials<sup>46,47</sup>. On the other hand, according to ref. 48, large impurity concentrations have an adverse effect on charge carrier transport: the mobility due to impurity scattering can theoretically be shown to vary as  $\sim 1/N_T$ , where  $N_T$  is the total impurity concentration. Indeed, we measure smaller photocurrent efficiencies for larger-diameter nanorods (Supplementary Fig. 19b), for which the electrons have to travel longer distances to reach the ITO electrode.

For structural defects, high-resolution TEM shows that these nanorods have a structurally irregular surface layer (Extended Data Fig. 6g), even though the bulk of a nanorod is a single crystal (Extended Data Fig. 6h). There is also some indication of an irregular surface structure in the study by Liu *et al.*<sup>14</sup>, whose synthesis procedure we followed. The irregular surface structure could be caused by surface reconstruction of the rutile {100} surface facets, which was previously observed for rutile {100} single-crystal surfaces annealed in the presence of O<sub>2</sub><sup>49,50</sup> (our nanorod sample was synthesized and annealed in air). The reconstructed surface is oxygen deficient (that is, oxygen vacancies are present) and contains small Ti<sub>2</sub>O<sub>3</sub> units, which could behave as catalytic units for water oxidation<sup>15</sup>.

**The two probe molecules adsorb to all surface sites equivalently.** To test whether our two structurally similar probe molecules amplex red and resazurin sample all surface sites in an unbiased way, we analysed the residence time of the reaction product resorufin (which is also structurally similar to amplex red and resazurin; Extended Data Fig. 1a) on the nanorod as a function of position. Using the three nanorods we presented as examples, we separate each nanorod into 150-nm segments (for example, Extended Data Fig. 7a, b, d, e, g and h), and within each segment we calculated the average residence time of the individual resorufin molecules; this average residence time is a good measure of how strongly resorufin, and presumably amplex red and resazurin, adsorb on the surface. Across the different segments within each nanorod, the average residence time of resorufin is essentially equivalent ( $\sim 15$  ms) and is independent of location (Extended Data Fig. 7c, f, i), even though different segments differ greatly in their reaction rates. This independence of location for the average residence time indicates that resorufin, and presumably amplex red and resazurin, adsorb equally on all surface sites along the side facets. We also determined the average residence times of resorufin from the 78 nanoscale spots where  $v_{\text{AR}}-E$ ,  $v_{\text{RZ}}-E$  and  $i-E$  behaviours were analysed. The average product residence time is independent of the effective rate constants  $k_h$  and  $k_e$  (Extended Data Fig. 7j), further confirming that resorufin, and presumably amplex red and resazurin, do not preferentially adsorb onto higher- or lower-activity sites.

**Ensemble photoelectrochemical measurements.** Ensemble-level photoelectrochemical measurements were performed using a potentiostat connected to the three-electrode electrochemical microfluidic cell mounted on the optical microscope. Either a chopped 365-nm light-emitting diode (LED) or a 375-nm laser illuminated the sample. Two illumination conditions were used for the 375-nm laser excitation: (1) large illumination spot ( $1 \times 1 \text{ cm}^2$ ) by removing the objective and (2) prism-type total internal reflection illumination with a small spot size of  $80 \times 95 \mu\text{m}^2$ , so as to access a large range of power density of a few milliwatts per square centimetre up to  $0.2 \text{ kW cm}^{-2}$ .

**Sub-nanorod photocurrent measurements.** A focused 375-nm laser was used to excite individual spots on individual nanorods in a front-side illumination geometry through the electrolyte (Fig. 1a), with a focus diameter of 392 nm (Supplementary Fig. 4) and maximal power density of  $6.6 \text{ MW cm}^{-2}$ . This spatially localized carrier excitation ensured that the photocurrent response stemmed from a particular nanorod, even though there were many other nanorods on the same ITO electrode. The 375-nm laser excitation light was chopped at 1 Hz and a lock-in amplifier was used to detect the nanoampere-level current from a single nanorod on top of the microampere-level background current of the ITO electrode via an analogue potentiostat (Supplementary Note 1.5). The  $i-E$  data of all individual nanorod spots were acquired in  $-0.1$  V increments from  $+0.2$  V to  $-0.7$  V.

**Quantitative analysis of  $i-E$ ,  $v_{\text{AR}}-E$  and  $v_{\text{RZ}}-E$  data.** We used a modified Gärtner–Butler model<sup>20,21</sup> to analyse the  $i-E$  data. This model assumes that the applied  $E$  is manifested entirely as a potential drop within the depletion layer of the semiconductor (that is, the depletion approximation), whose thickness  $W$  increases with  $E^{1/2}$ , and only charge carriers photogenerated within  $W$  contribute to the photocurrent.

Using a factor  $\eta$  as the absorbed-photon-to-current efficiency within  $W$ , we derived (Supplementary Note 3):

$$i = \eta q I_0 \alpha \left( \frac{s}{S} \right) W = \eta q I_0 \alpha \left( \frac{s}{S} \right) \sqrt{\frac{2 \varepsilon \varepsilon_0 (E - E_{\text{on,GB}})}{q N_d}} \quad (1)$$

This  $\eta$  contains contributions from both  $\eta_{\text{sep}}$ , the efficiency of charge carrier separation and transport to the collecting contact, and  $\eta_{\text{surf}}$ , the efficiency of charge transfer at the interface (Supplementary Note 5.3)<sup>23</sup>.  $q$  is the electronic charge,  $\alpha$  is the absorption coefficient at 375 nm ( $\alpha = 10^4 \text{ cm}^{-1}$ )<sup>51</sup>,  $s$  is the nanorod's geometric cross-section within the laser illumination area ( $S$ ),  $\varepsilon$  is the dielectric constant of rutile  $\text{TiO}_2$  ( $\varepsilon = 100$ )<sup>52</sup>,  $\varepsilon_0$  is the vacuum permittivity, and  $N_d$  is the doping density ( $N_d = 10^{18} \text{ cm}^{-3}$ ; see Supplementary Note 3.6).  $E_{\text{on,GB}}$  is the photocurrent onset potential predicted by the Gärtner–Butler model (that is, when  $E > E_{\text{on,GB}}$ ,  $i > 0$ ). For an ideal semiconductor–electrolyte interface (so that the depletion approximation applies) with minimal light absorption in the depletion region and negligible photocurrent contribution from carriers outside the depletion region, this  $E_{\text{on,GB}}$  is equivalent to the flat band potential for the semiconductor (see Supplementary Note 3.5). Possible nano-optical effects on the absorption coefficient  $\alpha$  of  $\text{TiO}_2$  nanorods as well as the changes caused by OEC deposition are considered in Supplementary Note 5.7; these effects do not affect any of our conclusions.

Equation (1) satisfactorily accounts for the dependence of  $i$  on  $I_0$  and  $i$  on  $E^{1/2}$  for  $E \geq -0.3 \text{ V}$  before and after OEC deposition (Fig. 2a, Extended Data Figs 4a, c and 8b–d). Data fitting gives  $\eta$  and  $E_{\text{on,GB}}$  for each spot of each nanorod before and after OEC deposition (Fig. 2a). Data at  $E \leq -0.4 \text{ V}$  were not fitted as they are complicated by photocurrent transient dynamics (Extended Data Fig. 4b).

The Gärtner–Butler model also allowed for treating the  $E^{1/2}$  and  $I_0^{1/2}$  dependences of  $v_{\text{AR}}$  by assuming that a primary step in water oxidation is hole transfer to surface-adsorbed  $\text{OH}^-$  or  $\text{H}_2\text{O}$  to generate  $\text{OH}^*$  radicals<sup>18,53</sup> and incorporating that amplex red reacts with  $\text{OH}^*$  radicals<sup>11,32,54</sup>. The  $v_{\text{Rz}}-E$  data could be treated by assuming that resazurin reacts with surface-accessible photogenerated electrons whose steady-state surface concentration is given by Reichman<sup>22</sup>. We derived (Supplementary Note 4):

$$v_{\text{AR}} = k_h \sqrt{I_0 \frac{\alpha s}{A S}} \sqrt{\frac{2 \varepsilon \varepsilon_0 (E - E_{\text{on,GB}})}{q N_d}} \quad (2)$$

$$v_{\text{Rz}} = \frac{k_e (I_0 / A) (s / S)}{\left( v_c + \mu_n \sqrt{\frac{2q}{\varepsilon \varepsilon_0} N_d (E - E_{\text{on,GB}})} \right) \left( 1 + \frac{1}{\alpha V_t} \sqrt{\frac{2q}{\varepsilon \varepsilon_0} N_d (E - E_{\text{on,GB}})} \right)} \quad (3)$$

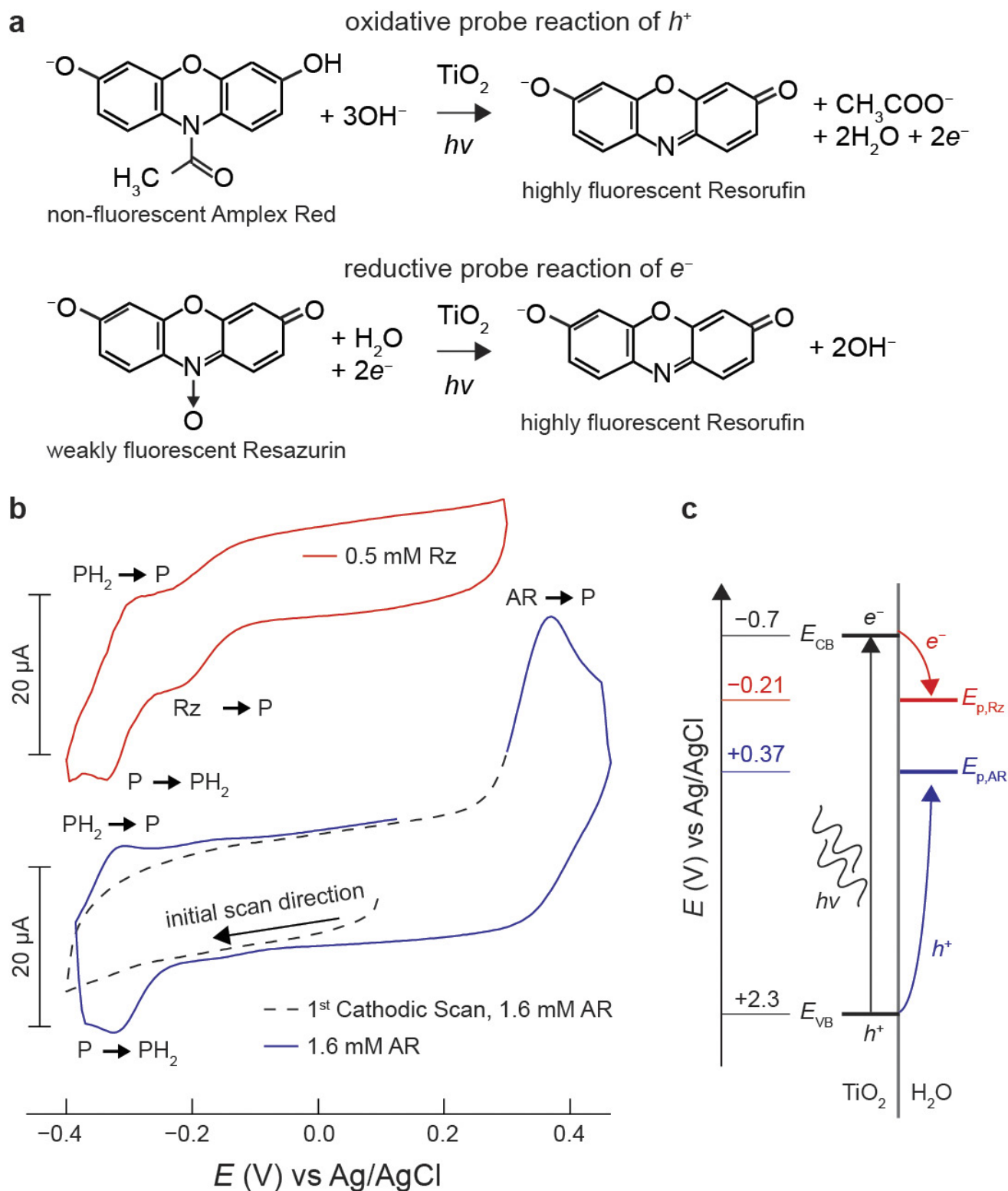
Here  $k_h$  and  $k_e$  are effective rate constants that reflect the hole- and electron-induced surface activities, respectively; both also depend on the respective (fixed) probe concentrations, and also implicitly have contributions from charge transport efficiencies to the nanorod surface.  $A$  is the electrochemically active surface area of the nanorod.  $v_c$ ,  $\mu_n$ , and  $V_t$  are the majority carrier recombination velocity ( $v_c = 3 \text{ cm s}^{-1}$ ; Supplementary Fig. 14), electron mobility ( $\mu_n \approx 10^{-4} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ )<sup>55</sup>, and thermal voltage ( $V_t \approx 0.026 \text{ V}$ ), respectively. Global fitting of  $v_{\text{AR}}-E$  and  $v_{\text{Rz}}-E$  data with equations (2) and (3) gave  $k_h$ ,  $k_e$  and  $E_{\text{on,GB}}$  for each spot of each nanorod where  $i-E$  data was also measured (Fig. 1l and m).

**Photoelectrochemical deposition and characterization of catalysts.** Individual spots of  $\text{TiO}_2$  nanorods were modified with a Co–Bi OEC<sup>56</sup> using a photoelectrochemical deposition method following previous studies of Co-based catalysts on metal oxide photoanodes<sup>30,57–60</sup>. Deaerated 1 M KCl, 100 mM sodium borate (pH 8.3) electrolyte containing 0.5 mM  $\text{CoCl}_2$  continuously flowed through the electrochemical cell while the potential was maintained at 0.0 V versus Ag/AgCl. Deposition occurred on specific locations of nanorods by focusing the 375-nm laser beam ( $552 \text{ kW cm}^{-2}$ ) onto a spot for 5 s (Extended Data Fig. 9a). A time-dependence experiment indicated that this deposition condition ensured that the photocurrent enhancement is saturated with the amount of deposited OEC (Extended Data Fig. 9b); this saturation behaviour is in agreement with the literature<sup>4,26</sup>. Consistently, the absorbed-photon-to-current efficiency  $\eta$  after OEC deposition is independent of the OEC coverage (Extended Data Fig. 9d). A representative optical transmission image of OEC-modified nanorods is shown in Extended Data Fig. 9a. The OEC appeared as a dark object on individual nanorods. The OEC coverage was estimated by analysing the SEM images of nanorods after OEC deposition (Supplementary Fig. 2). Here we first determined the bare nanorod structural contour in the SEM image. Then the area outside the contour and within the deposition laser focus was used as the measure for the amount of catalyst deposited in units of micrometres squared.

32. Xu, W., Jain, P. K., Beberwyck, B. J. & Alivisatos, A. P. Probing redox photocatalysis of trapped electrons and holes on single Sb-doped titania nanorod surfaces. *J. Am. Chem. Soc.* **134**, 3946–3949 (2012).
33. Chen, T.-Y. et al. Concentration and chromosome-organization dependent regulator unbinding from DNA for transcription regulation in living *Escherichia coli* cells. *Nature Commun.* **6**, 7445 (2015).
34. Sambur, J. B. & Chen, P. Approaches to single-nanoparticle catalysis. *Annu. Rev. Phys. Chem.* **65**, 395–422 (2014).
35. Byers, J. C., Güell, A. G. & Unwin, P. R. Nanoscale electrocatalysis: visualizing oxygen reduction at pristine, kinked, and oxidized sites on individual carbon nanotubes. *J. Am. Chem. Soc.* **136**, 11252–11255 (2014).
36. Chen, S. & Kucernak, A. Electrocatalysis under conditions of high mass transport rate: oxygen reduction on single submicrometer-sized Pt particles supported on carbon. *J. Phys. Chem. B* **108**, 3262–3276 (2004).
37. Meier, J., Friedrich, K. A. & Stimming, U. Novel method for the investigation of single nanoparticle reactivity. *Faraday Discuss.* **121**, 365–372 (2002).
38. Tel-Vered, R. & Bard, A. J. Generation and detection of single metal nanoparticles using scanning electrochemical microscopy techniques. *J. Phys. Chem. B* **110**, 25279–25287 (2006).
39. Ebejer, N., Schnipper, M., Colburn, A. W., Edwards, M. A. & Unwin, P. R. Localized high resolution electrochemistry and multifunctional imaging: scanning electrochemical cell microscopy. *Anal. Chem.* **82**, 9141–9145 (2010).
40. Li, Y., Cox, J. T. & Zhang, B. Electrochemical responses and electrocatalysis at single Au nanoparticles. *J. Am. Chem. Soc.* **132**, 3047–3054 (2010).
41. Xiao, X. & Bard, A. J. Observing single nanoparticle collisions at an ultramicroelectrode by electrocatalytic amplification. *J. Am. Chem. Soc.* **129**, 9610–9612 (2007).
42. Novo, C., Funston, A. M. & Mulvaney, P. Direct observation of chemical reactions on single gold nanocrystals using surface plasmon spectroscopy. *Nature Nanotechnol.* **3**, 598–602 (2008).
43. Shan, X. et al. Imaging the electrocatalytic activity of single nanoparticles. *Nature Nanotechnol.* **7**, 668–672 (2012).
44. Tang, M. L., Liu, N., Dionne, J. A. & Alivisatos, A. P. Observations of shape-dependent hydrogen uptake trajectories from single nanocrystals. *J. Am. Chem. Soc.* **133**, 13220–13223 (2011).
45. Plessers, E. et al. Resolving interparticle heterogeneities in composition and hydrogenation performance between individual supported silver on silica catalysts. *ACS Catal.* **5**, 6690–6695 (2015).
46. Xing, J. et al. Stable isolated metal atoms as active sites for photocatalytic hydrogen evolution. *Chem. Eur. J.* **20**, 2138–2144 (2014).
47. Ida, S., Kim, N., Ertekin, E., Takenaka, S. & Ishihara, T. Photocatalytic reaction centers in two-dimensional titanium oxide crystals. *J. Am. Chem. Soc.* **137**, 239–244 (2015).
48. Sze, S. M. *Semiconductor Devices, Physics and Technology* 3rd edn (Wiley, 2011).
49. Wang, Y., Warschkow, O. & Marks, L. D. Surface evolution of rutile  $\text{TiO}_2$  (100) in an oxidizing environment. *Surf. Sci.* **601**, 63–67 (2007).
50. Warschkow, O., Wang, Y., Subramanian, A., Asta, M. & Marks, L. D. Structure and local-equilibrium thermodynamics of the  $c(2 \times 2)$  reconstruction of rutile  $\text{TiO}_2$  (100). *Phys. Rev. Lett.* **100**, 086102 (2008).
51. Lindquist, S. E., Finnström, B. & Tegnér, L. Photoelectrochemical properties of polycrystalline  $\text{TiO}_2$  thin film electrodes on quartz substrates. *J. Electrochem. Soc.* **130**, 351–358 (1983).
52. Fàbrega, C. et al. Tuning the Fermi level and the kinetics of surface states of  $\text{TiO}_2$  nanorods by means of ammonia treatments. *J. Phys. Chem. C* **117**, 20517–20524 (2013).
53. Peterson, M. W., Turner, J. A. & Nozik, A. J. Mechanistic studies of the photocatalytic behavior of titania: particles in a photoelectrochemical slurry cell and the relevance to photodetoxification reactions. *J. Phys. Chem.* **95**, 221–225 (1991).
54. Shen, H., Zhou, X., Zou, N. & Chen, P. Single-molecule kinetics reveals a hidden surface reaction intermediate in single-nanoparticle catalysis. *J. Phys. Chem. C* **118**, 26902–26911 (2014).
55. Chen, H., Wei, Z., Yan, K., Bai, Y. & Yang, S. Unveiling two electron-transport modes in oxygen-deficient  $\text{TiO}_2$  nanowires and their influence on photoelectrochemical operation. *J. Phys. Chem. Lett.* **5**, 2890–2896 (2014).
56. Surendranath, Y., Dincă, M. & Nocera, D. G. Electrolyte-dependent electrosynthesis and activity of cobalt-based water oxidation catalysts. *J. Am. Chem. Soc.* **131**, 2615–2620 (2009).
57. Zhong, D. K., Cornuz, M., Sivula, K., Gratzel, M. & Gamelin, D. R. Photo-assisted electrodeposition of cobalt-phosphate (Co-Pi) catalyst on hematite photoanodes for solar water oxidation. *Energy Environ. Sci.* **4**, 1759–1764 (2011).
58. Ye, H., Park, H. S. & Bard, A. J. Screening of electrocatalysts for photoelectrochemical water oxidation on W-doped  $\text{BiVO}_4$  photocatalysts by scanning electrochemical microscopy. *J. Phys. Chem. C* **115**, 12464–12470 (2011).
59. Abdi, F. F. & van de Krol, R. Nature and light dependence of bulk recombination in Co-Pi-catalyzed  $\text{BiVO}_4$  photoanodes. *J. Phys. Chem. C* **116**, 9398–9404 (2012).
60. Pilli, S. K. et al. Efficient photoelectrochemical water oxidation over cobalt-phosphate (Co-Pi) catalyst modified  $\text{BiVO}_4$ /1D- $\text{WO}_3$  heterojunction electrodes. *Phys. Chem. Chem. Phys.* **15**, 14723–14728 (2013).
61. Çakir, S. & Arslan, E. Voltammetry of resazurin at a mercury electrode. *Chem. Pap.* **64**, 386–394 (2010).

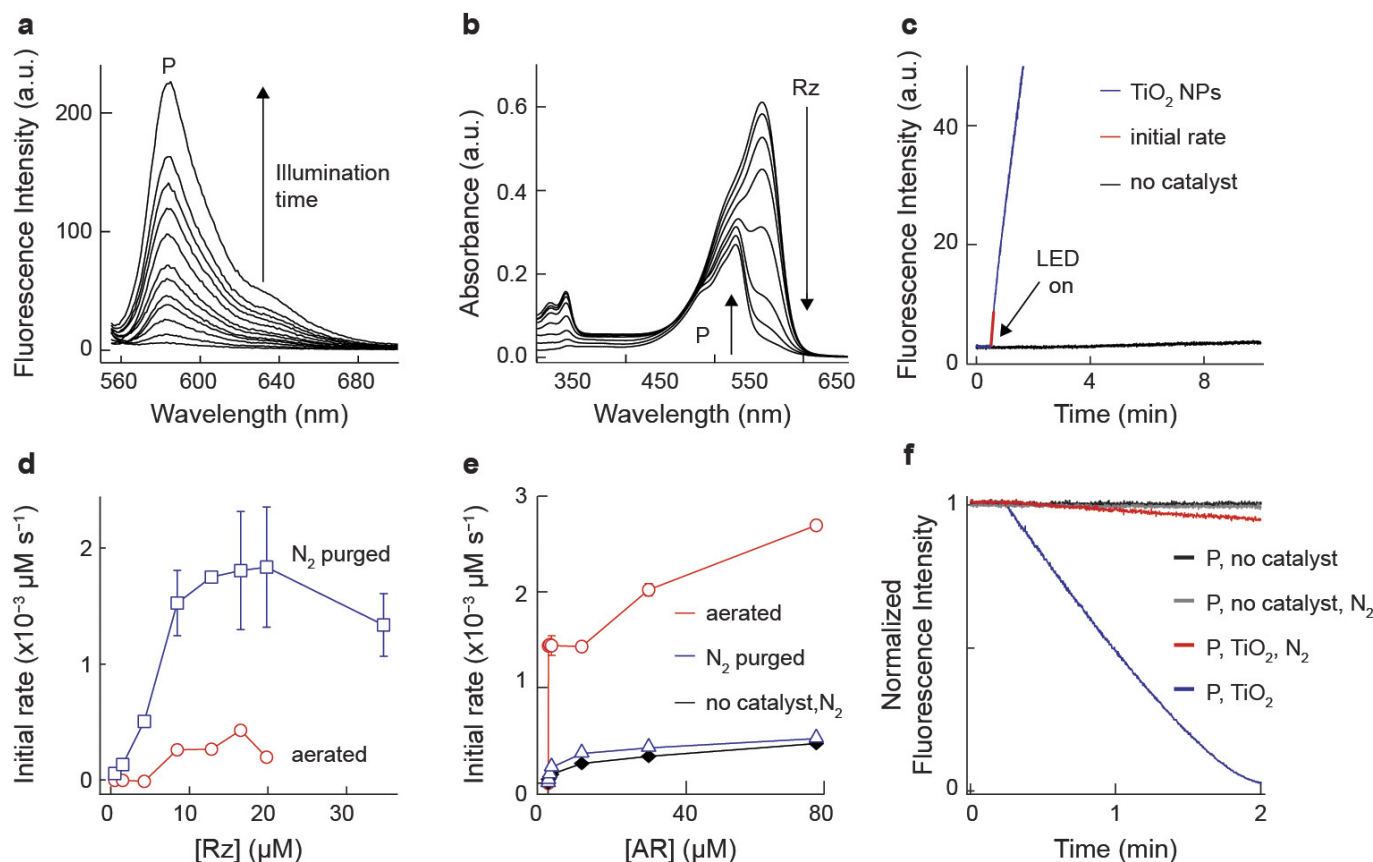


62. Cooper, G., Turner, J. A. & Nozik, A. J. Mott-Schottky plots and flatband potentials for single crystal rutile electrodes. *J. Electrochem. Soc.* **129**, 1973–1977 (1982).
63. Bolts, J. M. & Wrighton, M. S. Correlation of photocurrent-voltage curves with flat-band potential for stable photoelectrodes for the photoelectrolysis of water. *J. Phys. Chem.* **80**, 2641–2645 (1976).
64. Möllers, F., Tolle, H. J. & Memming, R. On the origin of the photocatalytic deposition of noble metals on TiO<sub>2</sub>. *J. Electrochem. Soc.* **121**, 1160–1167 (1974).
65. Laser, D. & Gottesfeld, S. Photocurrents induced by subbandgap illumination in a Ti-oxide film electrode. *J. Electrochem. Soc.* **126**, 475–478 (1979).
66. Abrantes, L. M. & Peter, L. M. Transient photocurrents at passive iron electrodes. *J. Electroanal. Chem. Interf. Electrochem.* **150**, 593–601 (1983).
67. Peter, L. M. Dynamic aspects of semiconductor photoelectrochemistry. *Chem. Rev.* **90**, 753–769 (1990).
68. Zhang, Q. *et al.* Density of deep trap states in oriented TiO<sub>2</sub> nanotube arrays. *J. Phys. Chem. C* **118**, 18207–18213 (2014).
69. Sagara, T. & Sukigara, M. Photogenerated surface states at n-TiO<sub>2</sub>/aqueous solution interface. *J. Electrochem. Soc.* **135**, 363–367 (1988).
70. Tafalla, D., Pujadas, M. & Salvador, P. Direct measurements of flat-band potential shifts under illumination of the semiconductor-electrolyte interface by electrolyte electroreflectance. *Surf. Sci.* **215**, 190–200 (1989).



**Extended Data Figure 1 | Carrier-selective fluorogenic probe reactions, and electrochemistry and energy-level alignment of the probe molecule redox potentials relative to TiO<sub>2</sub> band edges. a,** The oxidative N-deacetylation of amplex red to resorufin probes photogenerated holes (or consequent oxidizing species, such as surface-adsorbed hydroxyl radicals, OH<sup>•</sup> radicals) and the reductive N-deoxygenation of resazurin to resorufin probes photogenerated electrons. **b,** Cyclic voltammetry of 0.5 mM resazurin (Rz, red trace) and 1.6 mM amplex red (AR, black and blue traces) in N<sub>2</sub>-purged electrolyte. Scan rate is 2 mV s<sup>-1</sup> and the working electrode is glassy carbon. The irreversible reduction wave with a peak potential  $E_{p,Rz} = -0.21$  V is due to the reduction of resazurin to

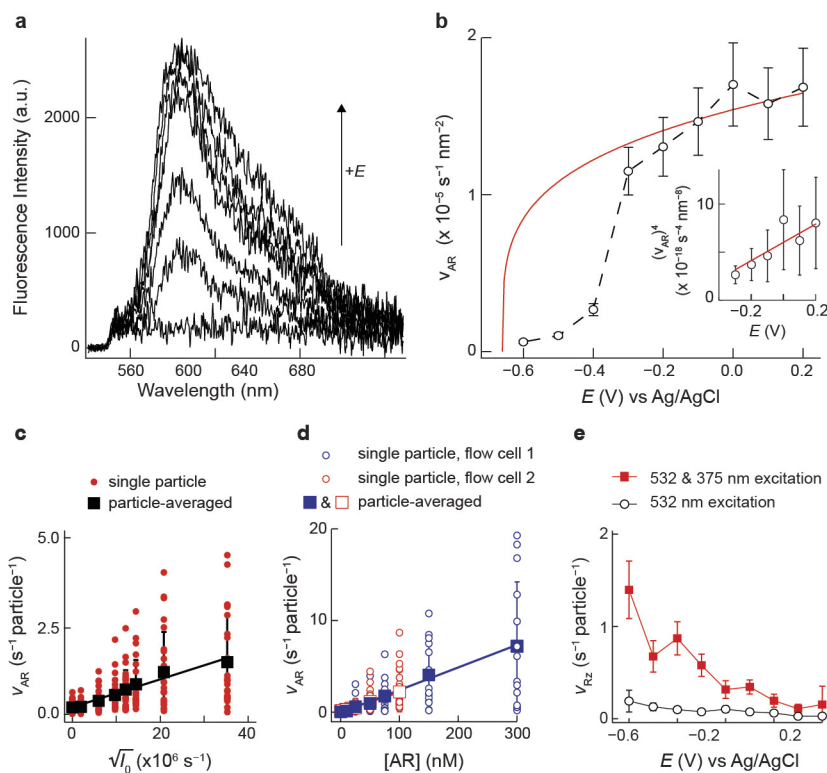
resorufin (product P)<sup>13,61</sup>. The reversible redox waves with a half-wave potential  $E_{1/2} = -0.32$  V are from the resorufin/dihydroresorufin redox couple (that is, resorufin/PH<sub>2</sub>)<sup>13,61</sup>. The irreversible oxidation wave was assigned to the irreversible oxidation of amplex red to resorufin, with a peak potential  $E_{p,AR} = +0.37$  V. **c,** Energy-level diagram of the peak potentials of resazurin reduction ( $E_{p,Rz}$ ) and amplex red ( $E_{p,AR}$ ) oxidation relative to the conduction ( $E_{CB}$ ) and valence band ( $E_{VB}$ ) edges of bulk TiO<sub>2</sub>. The blue and red arrows schematically illustrate the energetically possible pathways of photogenerated holes and electrons to amplex red and resazurin, respectively.  $h\nu$  indicates incident light.



**Extended Data Figure 2 | TiO<sub>2</sub> nanoparticles photocatalytically oxidize amplex red and reduce resazurin to generate resorufin.** Ensemble-level photocatalysis experiments were performed using P25 titanium dioxide nanoparticles, the precursor material for TiO<sub>2</sub> nanorods, because they were easily dispersed in aqueous solution. The nanorods were easily dispersed in ethanol–water mixtures, but this solvent is not ideal for studying amplex red oxidation because ethanol could preferentially react with photogenerated holes. **a** and **b**, Fluorescence and ultraviolet–visible absorption spectroscopy confirm the generation of resorufin from amplex red oxidation (**a**) or resazurin reduction (**b**) photo-catalysed by TiO<sub>2</sub> nanoparticles. **a**, Fluorescence spectra of a 10- $\mu\text{M}$  amplex red solution in an aerated pH 8.3 electrolyte as a function of 365-nm LED illumination time ( $45 \text{ mW cm}^{-2}$ ). **b**, Ultraviolet–visible absorption spectra of a 10- $\mu\text{M}$  resazurin solution in N<sub>2</sub>-purged pH 8.3 electrolyte as a function of illumination time; no isosbestic point is observed in the data, suggesting that resorufin is further consumed in the photocatalytic reaction. **c–e**, Photocatalytic kinetics of resazurin reduction (**c**, **d**) and amplex red oxidation (**e**) as a function of electrolyte composition. **c**, Fluorescence intensity at 585 nm versus time data used to determine the initial rate from a linear fit of the data after the 365 nm LED light is turned on (red line) for photocatalytic reduction of resazurin. The arbitrary fluorescence units were converted into concentration units by making a calibration curve using serial dilution of resorufin standards. NPs are nanoparticles. **d**, Initial rate of resazurin reduction versus bulk resazurin concentration at a fixed light intensity in aerated (red open circles) and N<sub>2</sub>-purged (blue open squares) electrolyte. The initial rate saturates with increasing resazurin concentration, [Rz], which suggests that the reaction occurs on

the catalyst surface and the number of available surface sites on the catalyst limits the initial reaction rate at high concentrations of the reactant resazurin. **e**, Initial rate of amplex red oxidation versus bulk amplex red concentration, [AR], in aerated (red open circles) and N<sub>2</sub>-purged (blue open triangles) electrolyte. A control experiment without catalyst is also shown (black filled diamonds). In N<sub>2</sub>-purged electrolyte the initial amplex red rate also exhibits saturation kinetics with increasing [AR], but the saturation level is much lower than that in aerated solution. This lower-saturation oxidation rate may be due to the removal of the possible O<sub>2</sub><sup>•−</sup>-induced amplex red oxidation channel or electron accumulation in TiO<sub>2</sub> particles in the absence of electron acceptors in the solution (for example, dissolved O<sub>2</sub>), which will not occur in photoelectrocatalysis experiments because electrons will be collected by the working electrode. The autocatalytic amplex red conversion to resorufin under ultraviolet illumination does not contribute much to our single-molecule imaging results of photoelectrocatalytic oxidation of amplex red on TiO<sub>2</sub> because the process occurs predominantly in the bulk solution rather than on the surface of TiO<sub>2</sub> and would not depend on the electrode potential. **f**, Decomposition of resorufin by TiO<sub>2</sub> nanoparticles in aerated (blue line) and N<sub>2</sub>-purged electrolyte (red line). This process could contribute to the short resorufin residence time on the TiO<sub>2</sub> surface (Extended Data Fig. 7). The rate of resorufin decomposition is accelerated in aerated electrolyte, similar to amplex red, which suggests an oxidative pathway for the decomposition of resorufin. The black and grey lines are controls. The reaction volume in each experiment was 2 ml, into which 100  $\mu\text{l}$  of a 1  $\text{mg ml}^{-1}$  aqueous solution of P25 nanoparticles was added.

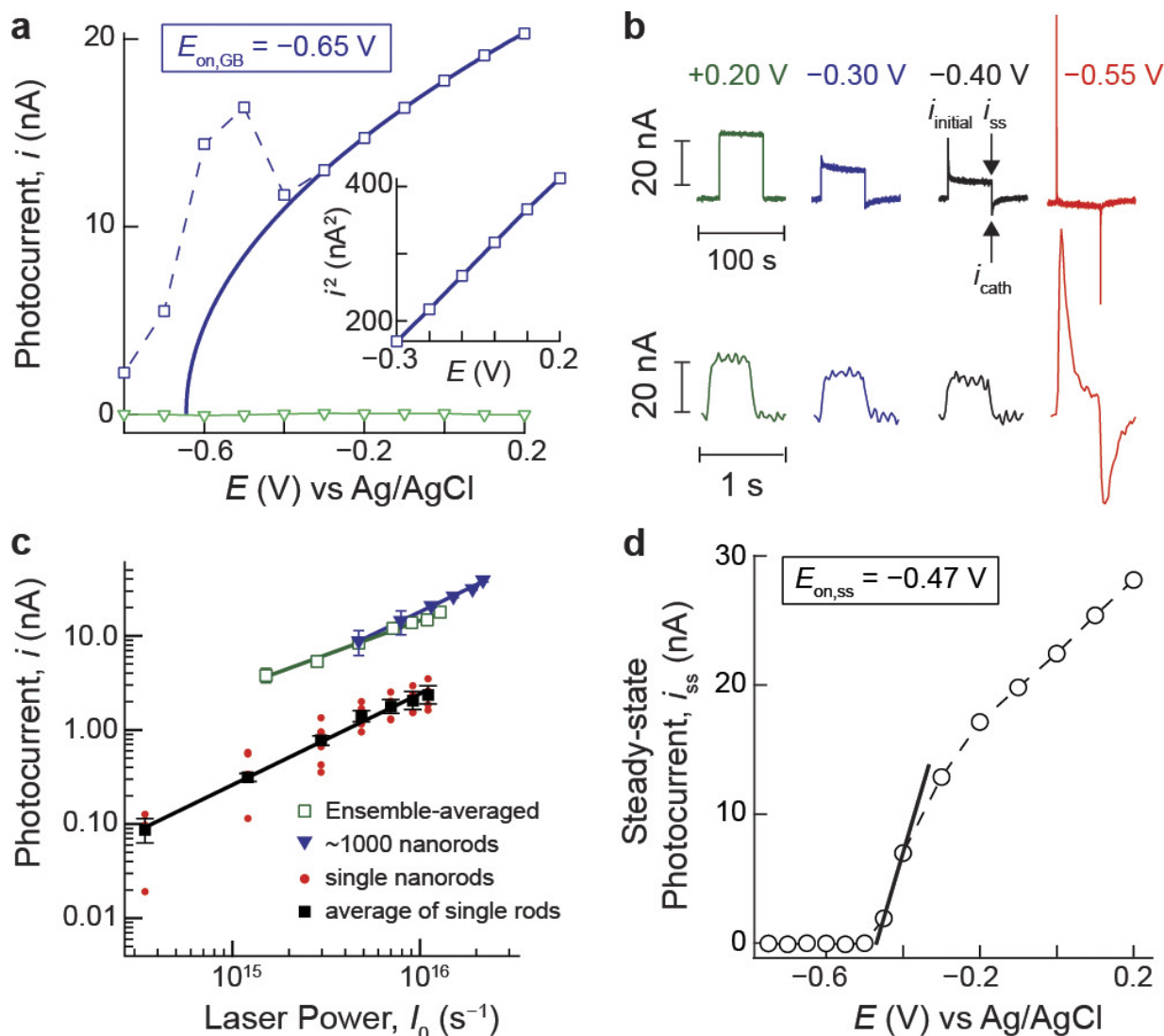




**Extended Data Figure 3 | TiO<sub>2</sub> nanorods photoelectrocatalytically oxidize amplex red and reduce resazurin to generate resorufin.**

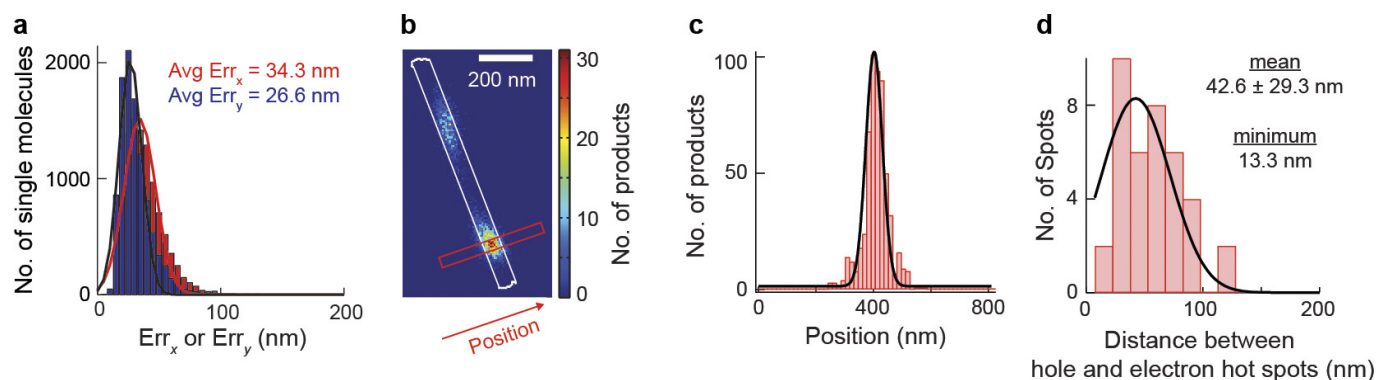
**a–d,**  $v_{AR}$  scales with  $E^{1/4}$  for  $E \geq -0.3$  V, with  $I_0^{1/2}$  at fixed positive  $E$ , and also scales with [AR] at fixed  $E$  and  $I_0$ . **a,** A series of fluorescence spectra measured from a single TiO<sub>2</sub> nanorod on ITO as a function of applied potential from  $-0.6$  V to  $+0.0$  V in N<sub>2</sub>-purged electrolyte with 50 nM amplex red under 375-nm laser illumination of  $60 \text{ W cm}^{-2}$  and 532-nm laser illumination of  $630 \text{ W cm}^{-2}$ . Each spectrum was acquired for 5 s, and it shows the characteristic fluorescence spectrum of resorufin. **b,** Single-molecule amplex red oxidation rate ( $v_{AR}$ ) averaged over the 37 nanorods used in the study. The red line represents a fit to the data with equation (2) in the main text to illustrate the rate dependence on the applied potential (an ensemble-averaged  $E_{\text{on,GB}} = -0.64$  V was obtained here). Only the data more positive than  $-0.3$  V were fitted, for the reasons discussed in Extended Data Fig. 4b caption. The error bars represent s.e.m. The inset plots  $v_{AR}^4$  versus  $E$  to show the linear behaviour at  $E$  more positive than  $-0.3$  V. **c,** Amplex red oxidation rate per particle versus the square root

of incident light power in the presence of 50-nM bulk [AR] and  $+0.2$  V, demonstrating that  $v_{AR}$  scales with  $I_0^{1/2}$  at a fixed positive  $E$ . The red dots indicate data from individual particles and the black squares represent the average (error bars represent s.d.). **d,** At  $+0.2$  V and 0.66-mW, 375-nm illumination (the highest light power in **c**) the reaction rate was observed to increase linearly with bulk [AR]. Two sets of data were included from two different flow cells, where the dots indicate individual particles and the solid squares represent the average (error bars are s.d.). Therefore, under our experimental single-molecule imaging conditions, the reaction rate is in the linear range in its dependence on [AR]. **e,** Resazurin is dominantly reduced by photogenerated conduction band electrons. Single-particle, single-molecule data of resazurin reduction averaged over a large number of TiO<sub>2</sub> nanorods in N<sub>2</sub>-purged electrolyte contained 50 nM resazurin. The control (black circles) without 0.66-mW, 375-nm laser illumination is also shown. The difference here demonstrates that resazurin is dominantly reduced by photogenerated conduction band electrons, while resazurin reduction under dark is negligible.



**Extended Data Figure 4 | Photoelectrochemical water oxidation properties of TiO<sub>2</sub> nanorods.**  $i$  scales with  $E^{1/2}$  for  $E \geq -0.3$  V and with  $I_0$  at fixed  $E$ , and determination of  $E_{\text{on,ss}}$ . **a**,  $i$ - $E$  data of ~1,000 nanorods measured with lock-in detection method using 1-Hz chopped 20-mW, 532-nm light illumination (green triangles) and 12-mW, 375-nm laser illumination (blue squares) in the TIRF geometry, demonstrating that anodic photocurrent was only observed under band gap illumination with the 375-nm laser. The solid line is a fit to the data at  $E \geq -0.3$  V with equation (1) in the main text to yield  $E_{\text{on,GB}} = -0.65$  V. This  $E_{\text{on,GB}}$  value agrees with the photocurrent onset potential determined for bulk rutile TiO<sub>2</sub>, which is also nearly equivalent to the flat band potential determined under dark conditions<sup>62,63</sup>. The inset shows that  $i^2$  scales linearly with  $E$  at  $E \geq -0.3$  V. For  $E < -0.3$  V, the photocurrent deviates from the  $E^{1/2}$  dependence owing to photocurrent transient dynamics (see **b**), which have been previously observed and described in detail<sup>19,29,64–68</sup>. **b**, Representative photocurrent-time ( $i$ - $t$ ) responses over a range of potentials during a single on-off cycle at 10 mHz (top) or 1 Hz (bottom) light chopping (50% duty cycle) to mimic steady-state single-molecule imaging and single-nanorod photocurrent measurement conditions, respectively. The background dark current measured from the macroscopic ITO electrode was subtracted using a two-point linear background subtraction method to an average of five consecutive data points before and after illumination. Data was collected with 20-ms time resolution. Three distinct features are clear (denoted on the -0.4-V data; black trace): (1) an initial photocurrent spike ( $i_{\text{initial}}$ ) when the light is turned on, (2) a decay of the photocurrent on a millisecond-to-second timescale while the light is on towards a steady-state current ( $i_{\text{ss}}$ ), and (3) an initial negative, cathodic photocurrent spike ( $i_{\text{cath}}$ ) when the illumination is turned off, which decays to the

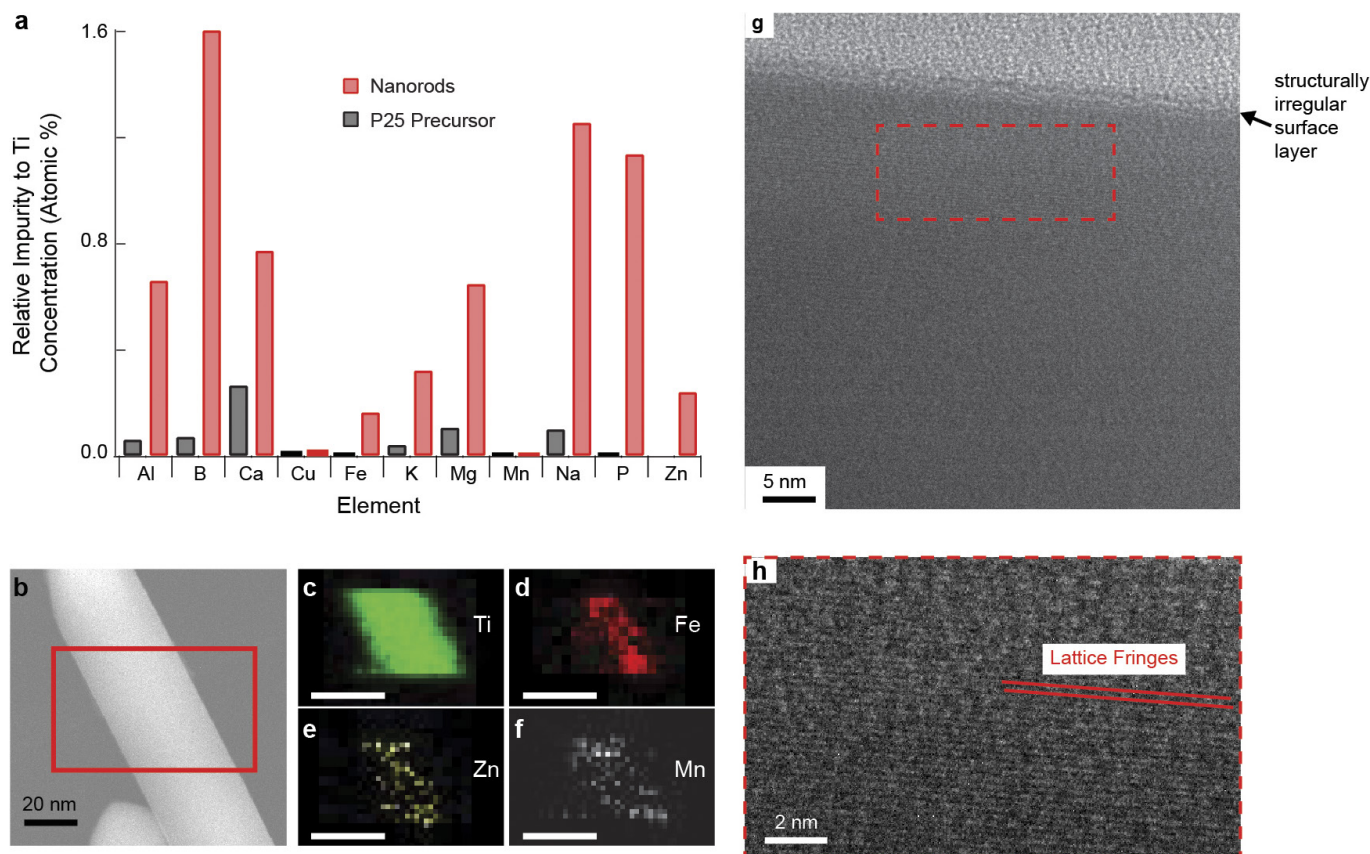
background current level also on a millisecond-to-second timescale. When the photocurrent transient behaviour dominates the  $i$ - $t$  response at  $E < -0.3$  V, the photocurrent signal measured by the lock-in amplifier is somewhere between the initial photocurrent spike ( $i_{\text{initial}}$ ) and steady-state photocurrent ( $i_{\text{ss}}$ ), which is expected because the lock-in detection is dependent on the signal shape versus time. Following references 19 and 69, we fit  $i$ - $E$  data for  $E \geq -0.3$  V in **a**. **c**, Photocurrent versus 375-nm laser power  $I_0$  at +0.2 V for three different illumination conditions, demonstrating that at fixed positive potentials,  $i$  scales linearly with  $I_0$  over a broad range of power densities (milliwatts to megawatts per centimetre squared). The geometric spot size of the laser spot was (1)  $1 \times 1$  cm<sup>2</sup> for ensemble-averaged measurements (open green squares, spanning a power density range of 0.8–6.8 mW cm<sup>-2</sup>), (2)  $80 \times 95$   $\mu$ m<sup>2</sup> to excite ~1,000 nanorods in the TIRF geometry (solid blue triangles, spanning a power density range of 32.9–151.3 W cm<sup>-2</sup>), and (3)  $390 \times 390$  nm<sup>2</sup> for single-nanorod measurements (solid red circles, spanning a power density range of 0.1–6.6 MW cm<sup>-2</sup>). **d**, Steady-state photocurrent  $i_{\text{ss}}$ - $E$  data (black circles, same data as included in Fig. 1n) obtained from 10-mHz chopped-light illumination experiments. The steady-state photocurrent onset potential ( $E_{\text{on,ss}}$ ) is defined as the intersection point of zero photocurrent and the tangent at maximum slope of photocurrent<sup>57</sup>. The solid black line indicates the linear fit to determine  $E_{\text{on,ss}} = -0.47$  V. This  $E_{\text{on,ss}}$  value is ~200 mV more positive than the  $E_{\text{on,GB}}$  value ( $-0.65$  V) in **a**; this difference has been previously observed and attributed to surface charging<sup>19,70</sup>. Despite the difference in their absolute values, the relative values and variations of  $E_{\text{on,GB}}$  among individual nanorod spots are good reflections of the relative values and variations of their  $E_{\text{on,ss}}$ , as  $E_{\text{on,GB}}$  is linearly correlated with  $E_{\text{on,ss}}$  (see Supplementary Fig. 20b in Supplementary Note 5.4).



**Extended Data Figure 5 | Product localization accuracy, hotspot size, as well as position correlation of hole- and electron-induced reaction hotspots.** **a**, Localization precision of product molecule positions. Distribution of the error in  $x$  and  $y$  (that is,  $\text{Err}_x$  or  $\text{Err}_y$ ) for all single molecules detected on the surface of a single nanorod, calculated using Supplementary equation (3). The average errors were determined with a one-dimensional Gaussian fit (solid lines). **b–d**, Size of localized reaction hotspots, and distribution of distances between corresponding high-reaction-rate electron and hole hotspots. **b**, Two-dimensional histogram ( $15 \times 15 \text{ nm}^2$  bins) of all hole-induced reaction product positions over the potential range  $E = -0.6 \text{ V}$  to  $+0.2 \text{ V}$  of a single

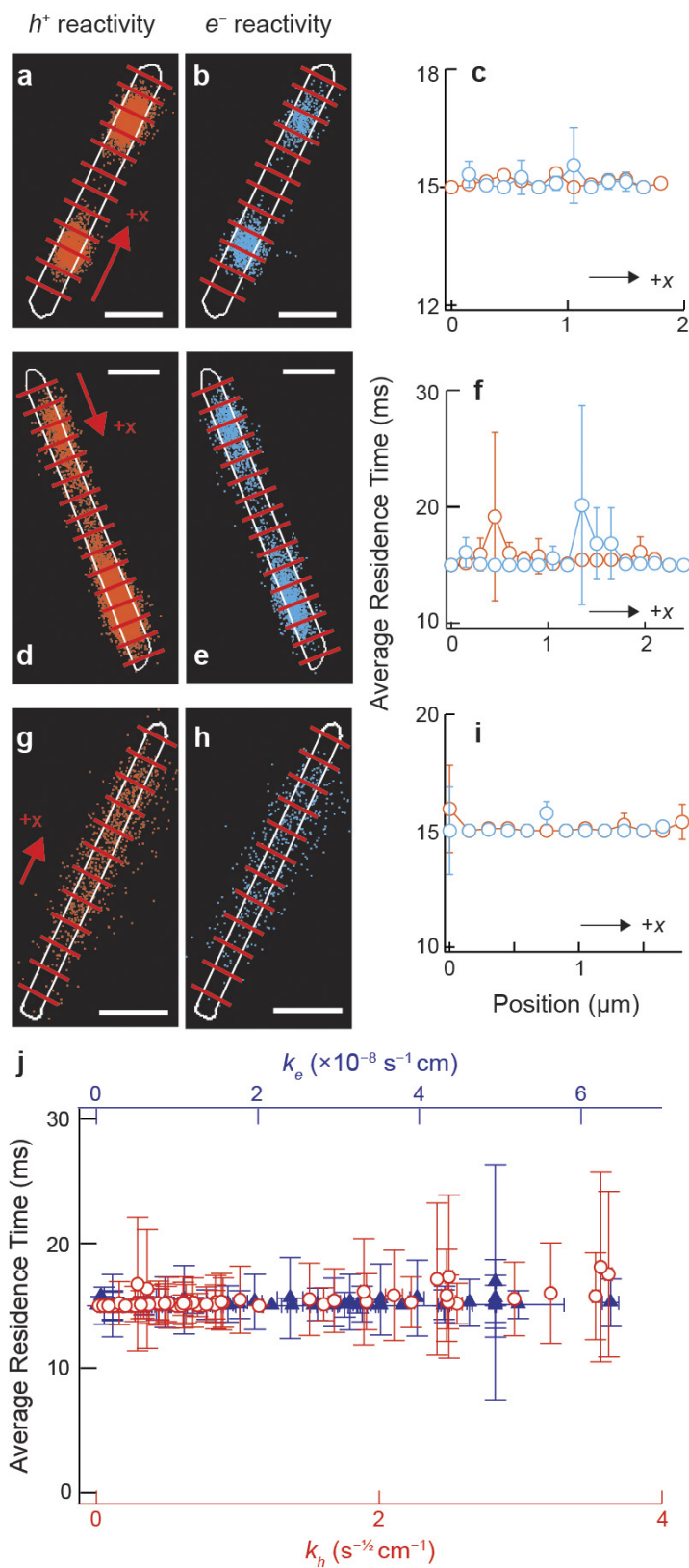
nanorod exhibiting a localized, hole activity hotspot. The red rectangular area represents the region where a one-dimensional histogram was plotted in **c**. **c**, One-dimensional histogram from the region in **b**. The solid line is a Gaussian fit. The full-width at half-maximum (FWHM), which is used as a measure of this hotspot size, is  $52 \text{ nm}$ , much smaller than the diffraction-limited resolution of  $\sim 300 \text{ nm}$ . **d**, Distributions of distances between corresponding hole and electron reaction hotspots on individual nanorods. Each hotspot's centre position was determined by two-dimensional Gaussian fitting to two-dimensional reaction product histograms. The average distance between localized hole and electron hot-spots is  $\sim 42 \text{ nm}$ , and the minimum distance is  $\sim 13 \text{ nm}$ .





**Extended Data Figure 6 | Elemental analysis and high-resolution electron microscopy for identifying possible active sites on these  $\text{TiO}_2$  nanorods.** See Methods for more details. **a**, Atomic concentration of impurity elements relative to Ti measured via ICP-AES in commercial P25 nanoparticles (black bars, precursor used in the synthesis) and the as-synthesized  $\text{TiO}_2$  nanorod sample (red bars). The dried powder samples were digested in 1 ml each of hot, concentrated  $\text{HNO}_3$  and  $\text{HCl}$  and submitted for elemental analysis at the Cornell Nutrient Analysis Laboratory. The total amount of  $\text{TiO}_2$  dissolved in the acid digestion process was difficult to quantify, so the concentration of impurity elements in each sample were calculated relative to the total amount of Ti detected. The ICP-AES data suggests that substantial concentration increases in elements such as Al, B, Ca, Fe, K and Mg, as well as the introduction of Zn, are due to impurity incorporation during the high-temperature synthetic

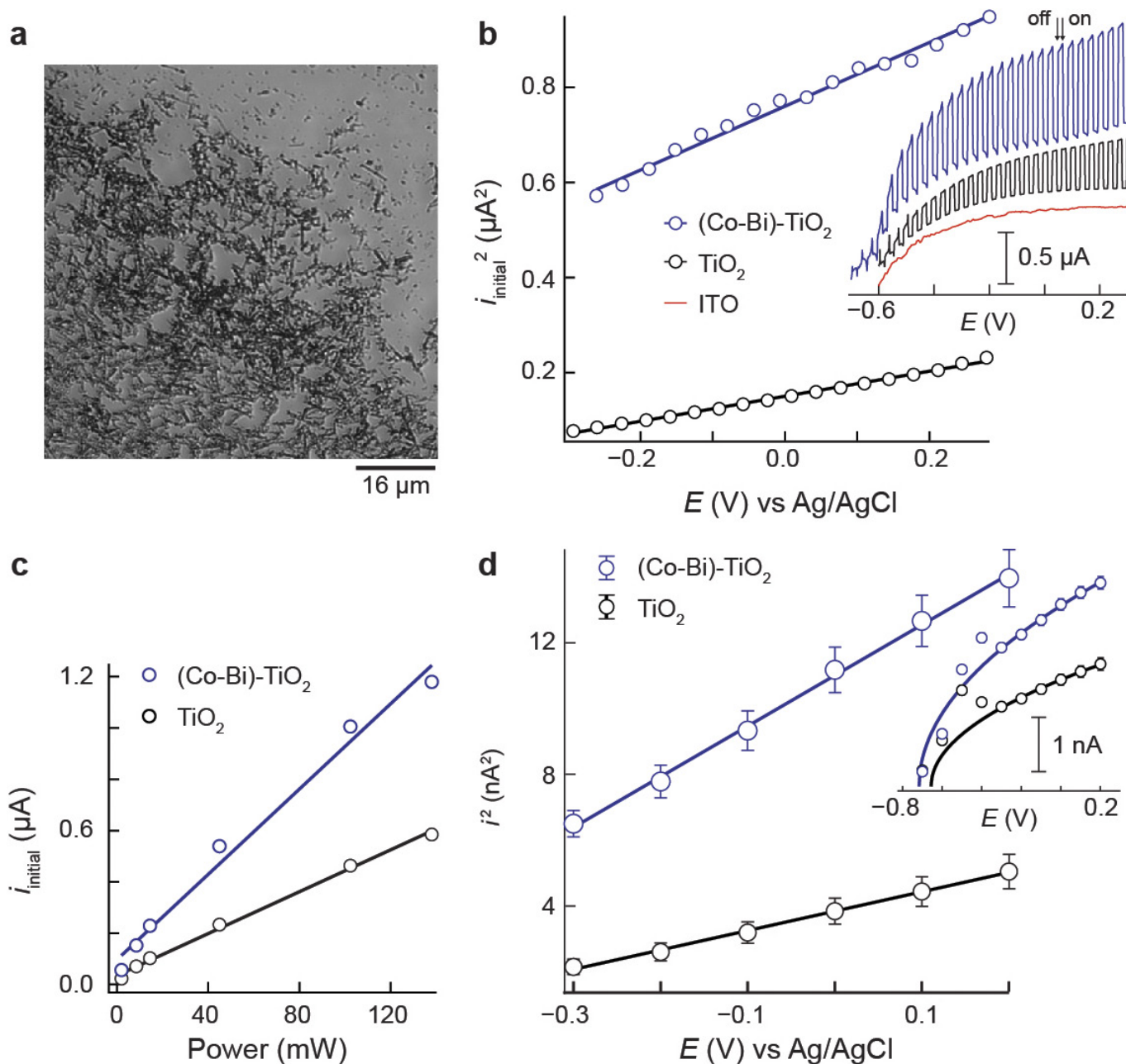
method, probably from reagent impurities (for example, in reagent-grade  $\text{NaCl}$  and  $\text{Na}_2\text{HPO}_4$ ). This hypothesis is further supported by ref. 14, which demonstrated that eight different transition metals (including Fe) could be substitutionally doped for Ti at an atomic concentration of 2%. **b**, High-angle annular dark field scanning TEM (HAADF-STEM) image of a single nanorod. The red box indicates a region of the image where energy-dispersive X-ray analysis was performed. **c–f**, Energy-dispersive X-ray analysis mapping of Ti and other detected impurity elements revealed a heterogeneous distribution of impurities: Fe (**d**), Zn (**e**) and Mn (**f**). **g**, High-resolution bright field TEM image of a single nanorod shows structural irregularities along the side surfaces. **h**, Zoomed-in view from the dashed red box in **g** shows lattice fringes, evidence that each rod is a single crystal. Similar lattice fringes were also observed by ref. 14.



**Extended Data Figure 7 | The two structurally similar probe molecules amplex red and resazurin adsorb to all surface sites equivalently.**

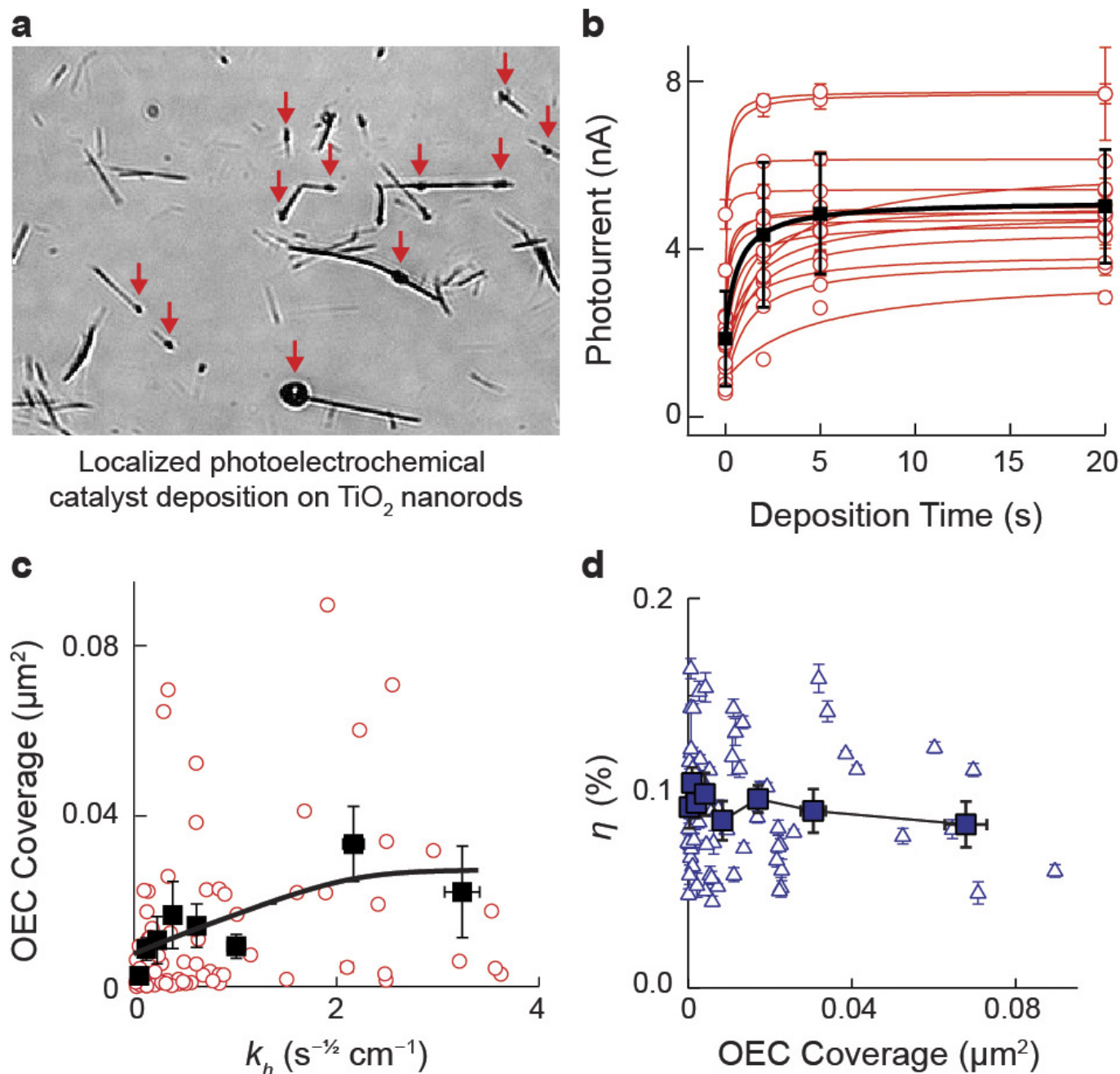
See Methods for more details. **a, b, d, e, g, h**, Scatter plots of all individual resorufin molecules generated from (**a, d and g**) hole-induced amplex red oxidation (orange dots) and (**b, e and h**) electron-induced resazurin reduction reactions (blue dots) on the same three  $\text{TiO}_2$  nanorods shown in Fig. 1. The red lines dissect each nanorod into 150-nm segments along its length, within each of which the average residence time is calculated.

**c, f and i**, Average product residence time versus position along the nanorod for hole-induced amplex red oxidation (orange circles) and electron-induced resazurin reduction (blue circles) reactions. **j**, Average product residence time formed from hole-induced amplex red oxidation (red circles) and electron-induced resazurin reduction (blue diamonds) reactions plotted versus the corresponding  $k_h$  and  $k_e$  values for all 78 nanorod spots studied herein. The error bars represent s.d.



**Extended Data Figure 8 | Photoelectrochemical properties of bare and (Co-Bi) modified TiO<sub>2</sub> nanorods.** **a–c**, Ensemble-level  $i$ - $E$  and  $i$ - $I_0$  data of bare and (Co-Bi)-modified TiO<sub>2</sub> nanorod thin films. **a**, Representative optical transmission image of a thin film of TiO<sub>2</sub> nanorods on an ITO electrode. **b**,  $i^2$  is linearly proportional to  $E$  for the same bare (black circles) and OEC-modified thin film sample as in **a** at  $E \geq -0.3$  V. Data were obtained from chopped light ( $138 \text{ mW cm}^{-2}$ , 365 nm LED) linear sweep voltammograms ( $-10 \text{ mV s}^{-1}$  from  $+0.3$  V, see inset) for a bare ITO electrode (red trace), TiO<sub>2</sub> nanorod-coated thin film electrode (black trace) and (Co-Bi)-TiO<sub>2</sub> electrode in N<sub>2</sub>-purged electrolyte (blue trace). The data have been slightly offset for clarity. Co-Bi OEC was photoelectrochemically deposited from electrolyte containing  $0.5 \text{ mM CoCl}_2$  at  $0.0$  V for  $30$  min using  $365\text{-nm}$  LED illumination at  $45 \text{ mW cm}^{-2}$ . **c**, Photocurrent is linearly proportional to  $365\text{-nm}$  light power at  $+0.3$  V for the same TiO<sub>2</sub> (red circles) and (Co-Bi)-TiO<sub>2</sub> (blue circles) thin film sample in N<sub>2</sub>-purged electrolyte. **d**, Single-nanorod

averaged  $i^2$  data are linearly proportional to  $E$  for bare (black circles) and Co-Bi OEC modified (red circles) TiO<sub>2</sub> nanorods at  $E \geq -0.3$  V. The data represents the average photocurrent of 78 nanorod spots from 37 individual nanorods and the error bars represent the standard error of the mean. The inset shows  $i$ - $E$  response obtained by lock-in detection under  $1\text{-Hz}$  illumination at the highest power density ( $6.6 \text{ mW cm}^{-2}$ ). The  $\eta$  and  $E_{\text{on,GB}}$  obtained from fitting the data (lines) with equation (1) in the main text are  $\eta = (5.72 \pm 0.11) \times 10^{-2}\%$  and  $E_{\text{on,GB}} = -0.66 \pm 0.02$  V for TiO<sub>2</sub>; and  $\eta = (9.62 \pm 0.17) \times 10^{-2}\%$  and  $E_{\text{on,GB}} = -0.72 \pm 0.02$  V for (Co-Bi)-TiO<sub>2</sub>. Together, these data demonstrate that the  $i$ - $E^{1/2}$  (at  $E \geq -0.3$  V) and  $i$ - $I_0$  scaling laws hold for both bare and OEC-modified nanorods, and in all cases, the photocurrent for the OEC-modified TiO<sub>2</sub> nanorod-coated electrode is higher than the bare TiO<sub>2</sub> nanorod-coated electrode. The OEC-induced photocurrent enhancement is also not due to larger photocurrent transient dynamics (Supplementary Note 5.7).

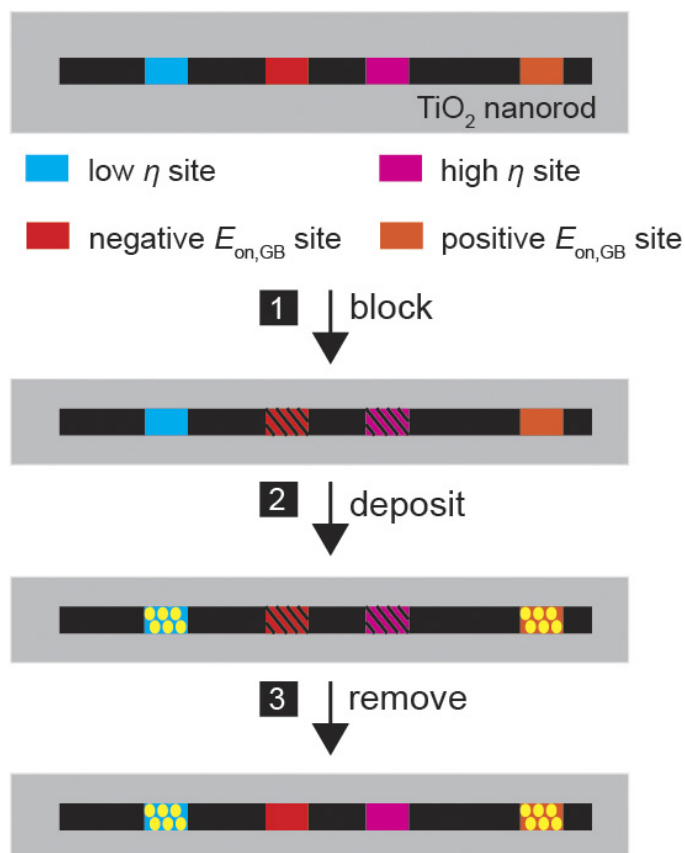


**Extended Data Figure 9 | Developing a Co-Bi OEC deposition method on single nanorods that produces a saturated photocurrent enhancement.** **a**, Representative optical transmission image of a nanorod sample after localized photoelectrochemical deposition of Co-Bi catalysts after 20 s deposition time. The red arrows indicate the location on which the focused laser was positioned for guided catalyst deposition, where the deposited catalysts appear as dark spots. **b**, Photocurrent at +0.2 V versus photoelectrochemical deposition time of the catalyst for single nanorods (red circles) and nanorod-averaged results (black squares). This photocurrent shows a saturation behaviour versus OEC deposition time (and hence OEC deposition amount), which was also observed for planar  $\text{Fe}_2\text{O}_3$  thin films modified with a Co-phosphate catalyst<sup>4,26</sup>. Error bars represent s.d. In all of our single-nanorod OEC deposition studies presented in the main text, we used an OEC deposition time of 5 s at which, according to **b**, the photocurrent enhancement is already saturated.

Thus, the  $\eta$  parameter obtained from fitting the  $i$ - $E$  data of OEC-modified nanorod spots represents the maximum achievable  $\eta$  per spot, as well as the maximum achievable change in  $\eta$  (that is,  $\Delta\eta$ ) between the same bare and OEC-modified nanorod spot. **c**, Catalyst coverage versus  $k_h$  (see Supplementary Fig. 2 for details on catalyst coverage quantification), demonstrating that more catalyst material is deposited onto higher-activity sites. The red circles represent data from individual nanorods and the black squares represent averaged data grouped by ten data points sorted by  $k_h$ . The same panel is also presented as Fig. 3c in the main text. **d**,  $\eta$  for OEC-modified nanorod spots versus catalyst coverage.  $\eta$  is essentially independent of OEC coverage, consistent with the saturated photocurrent behaviour in **b**. Filled squares represent data grouped by ten data points sorted by their catalyst coverage values. The error bars for all individual spot values represent s.d. The error bars in **c** and **d** for all binned data represent s.e.m.



## ITO electrode



**Extended Data Figure 10 | Proposed block-deposit-remove strategy, which has the potential to selectively deposit a desired OEC on low-activity and positive onset potential sites.** First, a removable protective group can be photoelectrochemically deposited starting from negative potentials to block negative  $E_{\text{on,GB}}$  sites. To block high-activity sites with similar  $E_{\text{on,GB}}$  values, we propose that using short light pulses may cause more protecting group material to be deposited at higher-activity sites first. In this way, low  $\eta$  and positive  $E_{\text{on,GB}}$  sites will remain unblocked. The potential could then be stepped from negative  $E$  to the ensemble-averaged steady-state photocurrent onset value, for example, and thus all remaining sites with  $E_{\text{on,GB}}$  values more positive than the ensemble-averaged value will also remain unblocked. Second, the potential can be set to a positive value and a long light pulse can be used to deposit the desired OEC on all remaining sites: those with positive  $E_{\text{on,GB}}$  and low  $\eta$ . Finally, the protective group can be subsequently removed to yield a photoanode with optimally located OECs. Yellow dots represent the OEC.

# Dehydration of lawsonite could directly trigger earthquakes in subducting oceanic crust

Keishi Okazaki<sup>1</sup> & Greg Hirth<sup>1</sup>

Intermediate-depth earthquakes in cold subduction zones are observed within the subducting oceanic crust, as well as the mantle<sup>1,2</sup>. In contrast, intermediate-depth earthquakes in hot subduction zones predominantly occur just below the Mohorovičić discontinuity<sup>1</sup>. These observations have stimulated interest in relationships between blueschist-facies metamorphism and seismicity, particularly through dehydration reactions involving the mineral lawsonite<sup>1,2</sup>. Here we conducted deformation experiments on lawsonite, while monitoring acoustic emissions, in a Griggs-type deformation apparatus. The temperature was increased above the thermal stability of lawsonite, while the sample was deforming, to test whether the lawsonite dehydration reaction induces unstable fault slip. In contrast to similar tests on antigorite, unstable fault slip (that is, stick-slip) occurred during dehydration reactions in the lawsonite and acoustic emission signals were continuously observed. Microstructural observations indicate that strain is highly localized along the fault (R<sub>1</sub> and B shears), and that the fault surface develops slickensides (very smooth fault surfaces polished by frictional sliding). The unloading slope during the unstable slip follows the stiffness of the apparatus at all experimental conditions, regardless of the strain rate and temperature ramping rate. A thermomechanical scaling factor<sup>3</sup> for the experiments is within the range estimated for natural subduction zones, indicating the potential for unstable frictional sliding within natural lawsonite layers.

To understand the spatial and temporal distribution of earthquakes in subduction zones, it is important to constrain the rheological properties of metamorphic rocks (that is, altered oceanic crust, mantle and sediments), and how they evolve during metamorphic reactions following hydration, carbonation and dehydration of the down-going slab. The difference between crustal seismicity in cold versus hot subduction zones (Fig. 1a) can be explained by differences in the thermal evolution of the slab and the location of major metamorphic reactions<sup>2,4</sup>. For example, metamorphosed oceanic crust in cold subduction zones (for example, northeastern Japan) is mainly composed of lawsonite-blueschist, whereas lawsonite-bearing rocks rarely occur in hot subduction zones (for example, Cascadia and southwestern Japan) because of a high thermal gradient (Fig. 1). Lawsonite, CaAl<sub>2</sub>Si<sub>2</sub>O<sub>7</sub>(OH)<sub>2</sub>·H<sub>2</sub>O, contains 11.5 wt% water in its crystal structure, and thus dehydration reactions involving lawsonite breakdown have been suggested to promote the upper plane of seismicity in cold subduction zones<sup>1,2,4,5</sup>. In some locations, crustal seismicity appears to be localized near the thermal stability limit of lawsonite<sup>6</sup>. In other locations, seismicity occurs over a broad range of conditions within the lawsonite-blueschist and lawsonite-eclogite facies, and then becomes rare below the transition to nominally anhydrous eclogite (Fig. 1b)<sup>1,2,5,7</sup>. Uncertainties in thermal models associated with the nature of decoupling along the slab–wedge interface make it somewhat difficult to constrain the exact relationship between dehydration reactions and seismicity. However, the continuous nature of dehydration reactions in altered basaltic rocks indicates

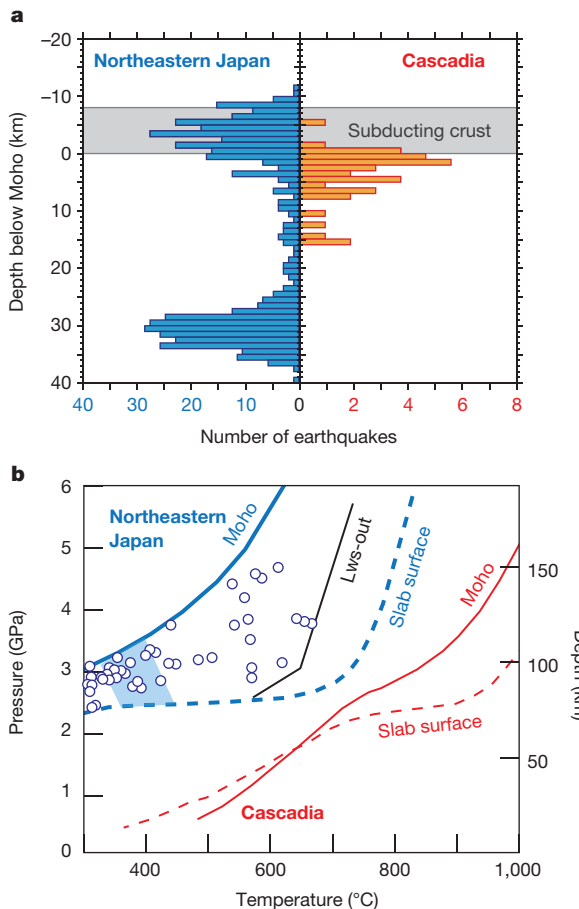
that lawsonite dehydrates over a wide range of conditions where these earthquakes are observed<sup>5,8</sup> (Fig. 1b).

We conducted triaxial deformation experiments on synthetic lawsonite using a Griggs-type deformation apparatus. The samples were hot-pressed before deformation and sheared between alumina pistons with surfaces oriented at 45° to the axial load. Acoustic emissions were monitored during the deformation experiments with a sampling rate of 2.5 MHz. Results were compared to experiments on antigorite serpentinite, for which stable frictional behaviour has been observed during dehydration at similar deformation conditions<sup>3,9,10</sup>. Deformation was initiated at a confining pressure of 1.0 GPa, temperatures of 300 °C (for lawsonite) and 400 °C (for antigorite), and constant displacement rates of 0.18–0.018 μm s<sup>−1</sup> (corresponding to equivalent strain rates  $\dot{\epsilon}$  of  $9 \times 10^{-5}$  s<sup>−1</sup> to  $9 \times 10^{-6}$  s<sup>−1</sup>; refs 11 and 12). Samples were first loaded at conditions within the stability field of lawsonite and antigorite to axial displacements of 0.9–1.3 mm (that is, for 2–2.5 h), at which point differential stresses  $\sigma_d$  were 2–2.5 GPa (Figs 2 and 3). Next the temperature was increased to 600 °C for lawsonite<sup>13</sup> and 700 °C for antigorite<sup>14</sup> to induce dehydration reactions while the samples continued to deform. The temperature ramping rates  $\dot{T}$  were varied from 0.5 °C s<sup>−1</sup> to 0.05 °C s<sup>−1</sup>. Samples were deformed for 6 min at the peak temperatures, then quenched.

Figure 2 shows the typical mechanical behaviour and cumulative acoustic emissions counts for temperature ramping experiments on lawsonite and antigorite. The result of a calibration test on pure aluminium, which deforms by dislocation creep at 300 °C and 1 GPa (ref. 15), is also shown; we used this test to assess the background acoustic emissions generated from cracking of the sample assembly (that is, graphite furnace and pyrophyllite) during the temperature ramping. Both lawsonite and antigorite show substantial weakening during temperature ramping; the lawsonite sample shows rapid slip coeval with a stress drop, while the antigorite sample shows relatively slow slip coeval with the stress drop. Increases in displacement rate (strain rate) reflect acceleration of the sample deformation resulting from elastic unloading of the column. Cumulative acoustic emissions for the lawsonite sample continuously increase at temperatures >450 °C (near the dehydration temperature of the lawsonite<sup>13,16</sup>), indicating unstable microcrack growth during fault slip. In contrast, the acoustic emissions signal for the antigorite sample is essentially the same as that observed for the aluminium sample, consistent with the lack of acoustic emissions observed in recent work on antigorite dehydration<sup>10</sup> (though also see previous work by ref. 17).

Rapid stress drops with unloading slopes ( $d\sigma_d/dx$ ) that are similar to the apparatus stiffness<sup>18</sup> indicate unstable slip during the lawsonite experiments. Thus, the use of a solid confining medium, which could inhibit rapid fault slip, did not suppress stick-slip behaviour (similar to previous observations in the Griggs apparatus<sup>19</sup>). During stable slip of the antigorite, the unloading slope is controlled by the temperature ramping rate and strain rate. Figure 3a and b shows differential stress and temperature as a function of axial displacement during the temperature ramping experiments. Figure 3c shows the unloading slope as

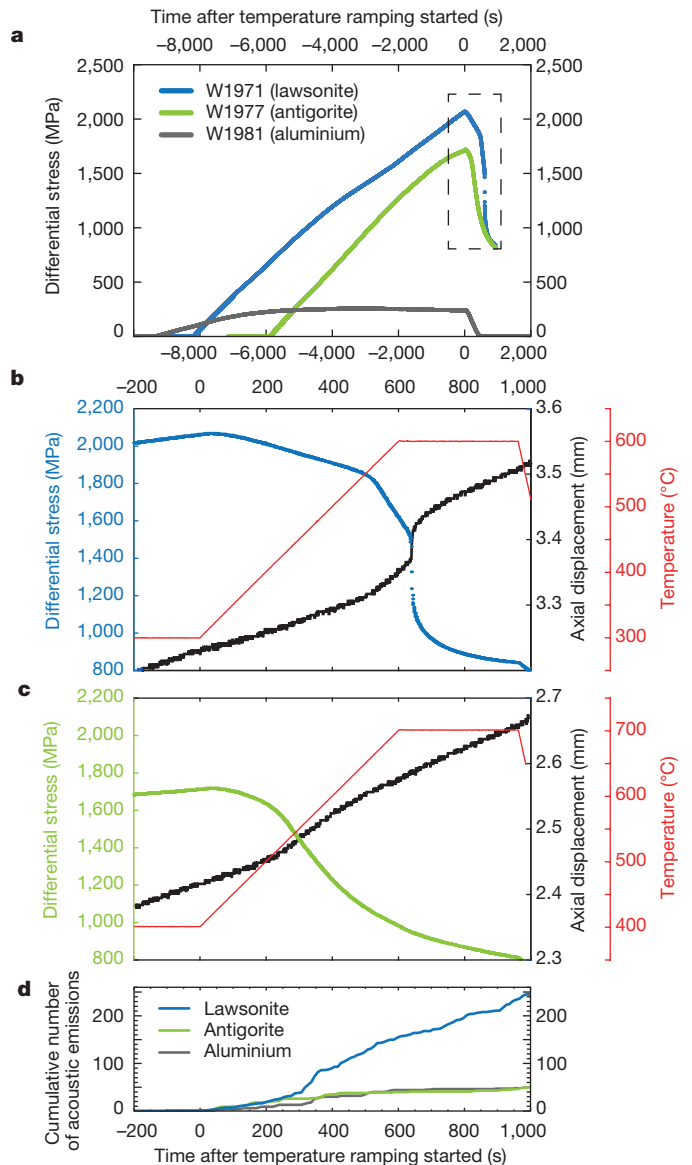
<sup>1</sup>Department of Earth, Environmental, and Planetary Sciences, Brown University, 324 Brook Street Box 1846, Providence, Rhode Island 02906, USA.



**Figure 1 | Distribution of intermediate-depth earthquakes, pressure-temperature paths for cold and hot subduction zones, and major dehydration reactions of lawsonite blueschist.** **a**, Number of earthquakes versus depth of subducting slabs<sup>2</sup>. The cold subduction zone (blue bars, Hokkaido in northeastern Japan) has a peak of seismicity within the subducting oceanic crust, as well as deeper in the subducting mantle (double seismic zone). The hot subduction zone (red bars, Cascadia in North America) has a single zone of seismicity near the Mohorovičić discontinuity (Moho), with little seismic activity within the subducting oceanic crust. **b**, Blue dots show the pressure-temperature conditions for which intermediate-depth earthquakes occur within the subducting oceanic crust in northern Japan. Coloured thick solid and dashed lines show pressure-temperature paths for the Moho and the slab surface of northeastern Japan (blue) and Cascadia (red), respectively<sup>2</sup>. The solid black line delineates the thermal stability limit of lawsonite (Lws-out) in the mid-ocean-ridge basaltic (MORB) system<sup>8,30</sup>. Dehydration reactions involving lawsonite in the MORB system occur continuously from  $\sim 300^\circ\text{C}$  to the stability limit<sup>5,8</sup>; the conditions at which these reactions initiate (which depend on bulk composition) are shown as the light blue field. Lawsonite is stable in cold subduction zones, where intermediate-depth earthquakes occur within the subducting oceanic crust, while zoisite is stable in hot subduction zones, where few intermediate-depth earthquakes are observed within the subducting oceanic crust.

a function of a time-independent thermomechanical factor  $\dot{T} / \dot{\epsilon}$ , the temperature ramping rate over the strain rate. The maximum unloading slope for lawsonite during the temperature ramping experiments is nearly constant and similar to the effective unloading stiffness of the apparatus ( $\sim 280 \text{ kN mm}^{-1}$ ,  $8.8 \text{ GPa mm}^{-1}$  for the sample dimension used in this study<sup>3</sup>) over the entire range of  $\dot{T} / \dot{\epsilon}$  tested. In contrast, the unloading slope for antigorite<sup>3</sup> decreases (becomes more negative) with increasing  $\dot{T} / \dot{\epsilon}$ . These observations indicate that dehydration induced unstable fault slip in lawsonite (that is, stick-slip) but stable slow slip in antigorite, consistent with the differences in acoustic emissions.

Recovered samples of lawsonite from temperature ramping experiments showed strong shear localization and evidence for dehydration



**Figure 2 | Differential stress, axial displacement and acoustic emissions counts as a function of time during temperature ramping experiments on lawsonite and antigorite.** **a**, Overall stress-time curves for lawsonite, antigorite and aluminium during temperature ramping experiments. Data were recorded at a sampling rate of 1 Hz, and all of these experiments were conducted at a confining pressure of 1 GPa, a displacement rate of  $0.18 \mu\text{m s}^{-1}$  (equivalent strain rate of  $9 \times 10^{-5} \text{ s}^{-1}$ ), and a temperature ramping rate of  $0.5^\circ\text{C s}^{-1}$ . **b** and **c**, Differential stress, axial displacement and temperature are plotted as dots as a function of time after temperature ramping initiates for experiments on lawsonite (**b**), and antigorite (**c**) (enlargements from the grey dashed rectangle **a**). **d**, Cumulative number of acoustic emissions as a function of time after initiation of temperature ramp.

reactions. Strain is highly localized along the  $R_1$  and B shear planes; the fault surface shows mirror-like slickensides with very little gouge (Fig. 4), similar to observations from high-pressure faulting associated with the olivine-spinel transition<sup>20</sup>. X-ray diffraction (XRD) analysis of the recovered samples indicates that a small amount of anorthite ( $< 5 \text{ vol}\%$ ) forms as a result of dehydration during the temperature ramping experiments (Fig. 4f and Extended Data Fig. 1). Similarly, chemical analysis of a lawsonite sample using a electron probe micro-analyser shows a slightly higher concentration of Si along the shear zone, subparallel to the slickenside surface; anorthite has a somewhat higher Si content ( $\sim 44 \text{ wt}\%$ ) than lawsonite ( $38.3 \text{ wt}\%$ ) and zoisite ( $\sim 39.7 \text{ wt}\%$ ).

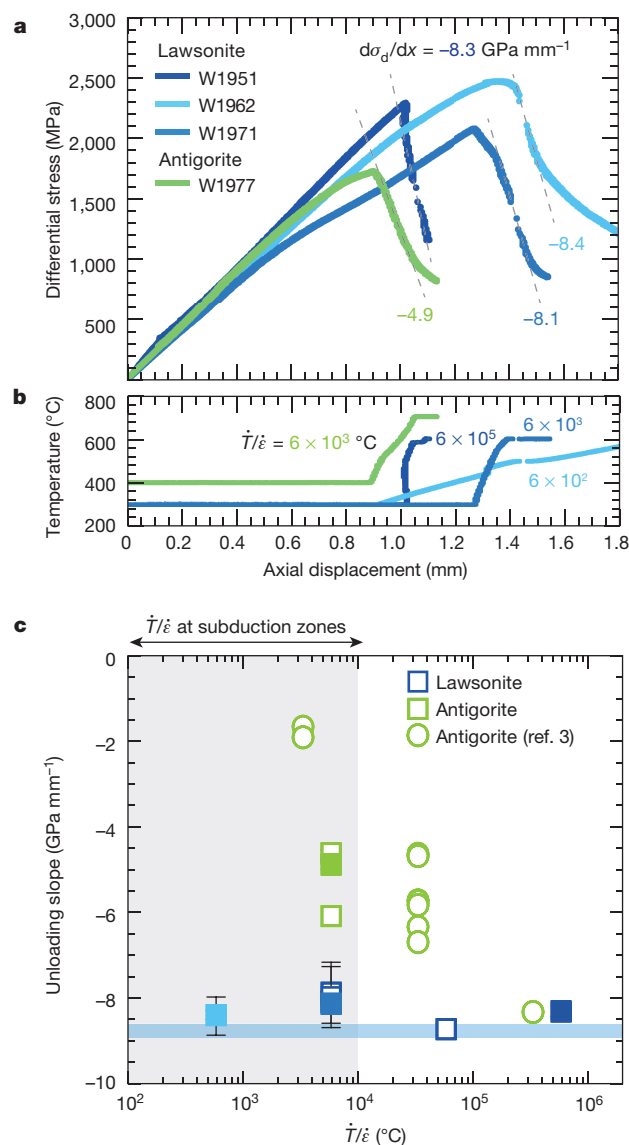
The difference in behaviour between antigorite and lawsonite during dehydration experiments mirrors differences in fault stability observed during experiments conducted within their respective stability fields. Experiments on antigorite show velocity-strengthening behaviour at the conditions of our experiments<sup>3,21</sup> at 400 °C. In contrast, we observe violent unstable stress drops during experiments on lawsonite at a constant temperature of 400 °C (see Extended Data Fig. 2), suggesting that lawsonite exhibits velocity-weakening friction—a property required for the development of unstable (seismic) fault slip<sup>22</sup>.

The application of our results to natural settings requires scaling in time and space. Scaling in time can be addressed using the time-independent thermomechanical factor  $\dot{T}/\dot{\epsilon}$ . While strain rates in our experiments are by necessity much greater than experienced inside the Earth, the values of  $\dot{T}/\dot{\epsilon}$  imposed in our experiments ( $6 \times 10^2$  °C to  $6 \times 10^5$  °C) cover the range expected for subduction zone systems (Fig. 3c). For example, with a thermal gradient of  $10^\circ\text{C km}^{-1}$  (ref. 23), a subduction velocity of  $10\text{ cm yr}^{-1}$ , a subduction angle of  $45^\circ$ , and a range of strain rate  $\dot{\epsilon}$  from  $10^{-13}\text{ s}^{-1}$  to  $10^{-15}\text{ s}^{-1}$ ,  $\dot{T}/\dot{\epsilon}$  in natural subduction zone systems ranges from  $\sim 10^2$  to  $10^4$  °C. The applicability of scaling by  $\dot{T}/\dot{\epsilon}$  is empirically supported by the observation of approximately constant unloading slopes in experiments on antigorite for which constant values of  $\dot{T}/\dot{\epsilon}$  are imposed with strain rates ranging from  $10^{-5}\text{ s}^{-1}$  to  $10^{-7}\text{ s}^{-1}$  (ref. 3). A comparison of experiments conducted under drained and undrained conditions indicates that the unloading slope for antigorite is primarily controlled by the effects of the dehydration reaction rate and strain rate on the evolution of pore fluid pressure<sup>3,24</sup>. Furthermore, deformation textures<sup>24</sup> and the observation of velocity-strengthening behaviour in partially reacted samples<sup>3</sup> suggest that the stable behaviour during dehydration of antigorite arises partly from the operation of dissolution–precipitation processes during semi-brittle flow. These observations indicate that kinetics of pore-fluid evolution resulting from reactions and deformation are balanced by varying  $\dot{T}$  and  $\dot{\epsilon}$  together. In contrast, the lack of scaling in  $\dot{T}/\dot{\epsilon}$  observed for the lawsonite experiments suggests that the kinetics of viscous deformation (that is, dissolution–precipitation processes) are too sluggish to prevent instability at the timescale of the dehydration reaction.

The potential stability of faults also depends on the effective stiffness of the system. The stiffness of natural fault zones  $K_f$  is approximated as:  $K_f = G/2(1 - \nu)/L$  (ref. 22), where  $G$  is the shear modulus ( $\sim 30\text{ GPa}$ ),  $\nu$  is the Poisson's ratio ( $\sim 0.25$ ), and  $L$  is the length of the slipping region. With these values, the stiffness of the natural systems ( $\sim 2\text{ MPa mm}^{-1}$  for  $L \approx 10\text{ m}$  (that is, for a M1 earthquake, which are prevalent in the analyses of refs 1, 2 and 7)) is orders of magnitude more compliant (less stiff) than that of the Griggs apparatus. Thus, unstable slip resulting from dehydration of lawsonite will occur even more easily in natural subduction zones than in the laboratory.

In contrast, the large difference in stiffness between the laboratory and natural settings does make it more difficult to predict the behaviour of antigorite-rich faults during dehydration. For example, the relationship between  $\dot{T}/\dot{\epsilon}$  and unloading slope (Fig. 3c) indicate that the unloading slope during reaction may be similar to that of the effective stiffness. Interestingly, slow earthquakes are observed in locations where antigorite dehydration is possible<sup>25</sup>. A variety of other processes have been proposed to explain intermediate-depth earthquakes in the subducting mantle, such as shear-heating instabilities promoted by localized deformation within fine-grained dehydration reaction products or gouge initially produced during outer-rise earthquakes<sup>26</sup>. In any case, the dramatic difference in the friction properties of lawsonite and antigorite during dehydration provides a new context in which to interpret the range of slip behaviours observed in natural settings.

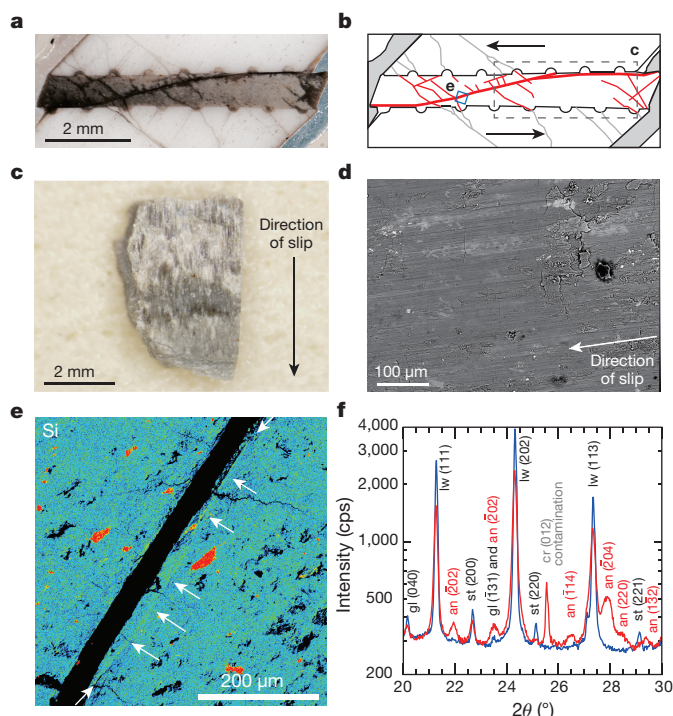
Our experiments demonstrate that lawsonite is one of the few minerals that exhibit brittle deformation resulting in unstable fault slip at the high pressure and temperature conditions appropriate for intermediate-depth earthquakes. In general, natural metamorphic rocks have a schistosity (layering structure such as foliation) and a very low permeability at high pressure ( $< 10^{-20}\text{ m}^2$ )<sup>27</sup>, so dehydration of lawsonite



**Figure 3 | Stress–displacement curves and unloading slopes during temperature ramping experiments on lawsonite and antigorite for a range of the thermomechanical factor  $\dot{T}/\dot{\epsilon}$ .** Differential stress (a), and temperature  $T$  (b) as a function of axial displacement during temperature ramping experiments on lawsonite (W1951, W1962 and W1971) and antigorite (W1977) at various ramping rates and strain rates. All data are plotted at a sampling rate of 1 Hz. Experiments are conducted at  $\dot{T}/\dot{\epsilon} \approx 6 \times 10^5$  (where  $\dot{T} = 0.5^\circ\text{C s}^{-1}$ ,  $\dot{\epsilon} \leq 9 \times 10^{-7}\text{ s}^{-1}$ ) for W1951,  $\dot{T}/\dot{\epsilon} = 6 \times 10^2$  ( $\dot{T} = 0.05^\circ\text{C s}^{-1}$ ,  $\dot{\epsilon} = 9 \times 10^{-5}\text{ s}^{-1}$ ) for W1962,  $\dot{T}/\dot{\epsilon} = 6 \times 10^3$  ( $\dot{T} = 0.5^\circ\text{C s}^{-1}$ ,  $\dot{\epsilon} = 9 \times 10^{-5}\text{ s}^{-1}$ ) for W1971, and  $\dot{T}/\dot{\epsilon} = 6 \times 10^3$  ( $\dot{T} = 0.5^\circ\text{C s}^{-1}$ ,  $\dot{\epsilon} = 9 \times 10^{-5}\text{ s}^{-1}$ ) for W1977, respectively. Grey dashed lines indicate unloading slopes ( $d\sigma_d/dx$ ) for each experiment. See Extended Data Table 1 for further details on experimental conditions. c, Unloading slopes during temperature ramping experiments. The light blue line shows the unloading stiffness of the apparatus<sup>3</sup>. Colours of filled symbols correspond to the colours of the lines in a and b. The unloading slopes of lawsonite samples are nearly constant at the unloading stiffness of the apparatus, whereas the other samples show how the unloading slope depends on  $\dot{T}/\dot{\epsilon}$ . Grey area indicates the expected range of  $\dot{T}/\dot{\epsilon}$  for natural subduction zones. Error bars represent 99% confidence intervals for the unloading slopes of each experiment.

could efficiently reduce the effective pressure and induce unstable slip with a reasonable stress drop along foliations in the lawsonite layers. Zones with high P-wave/S-wave velocity ratio are observed at the source region of intermediate-depth earthquakes within the subducting oceanic crust<sup>28</sup>, attesting to the presence of fluid. Lawsonite would





**Figure 4 | Lawsonite sample from a temperature ramping experiment.**

**a**, Photograph of the recovered sample (W1958). The section was cut parallel to the axial direction and the shear direction. **b**, Sketch of **a**; thick red lines indicate the major fault ( $R_1$  and B shear) formed during the temperature ramping experiment. Thin red lines show unloading cracks, which form perpendicularly to the axial direction. **c**, Section of fault surface (from the grey dashed rectangle in **b**) showing slickensides on the  $R_1$  and B shear plane. **d**, Back-scattered electron image of the surface shown in **c**. The dark grey matrix is lawsonite, medium light grey matrix is glaucophane, and light grey patches are staurolite. **e**, Si map of the region in **b** marked with a blue square. White arrows show the anorthite-rich (that is, Si-rich) layer subparallel to the fault surface shown in **c**. Higher-Si-content grains are glaucophane and staurolite. **f**, XRD spectra of the starting material (blue) and sample from experiment W1958 (red). Vertical and horizontal axes are the intensity in counts per second (cps) and the diffraction angle  $2\theta$ , respectively. Peaks for lawsonite (lw), glaucophane (gl) and staurolite (st) are present in the starting material; anorthite peaks (an) are present for W1958, indicating dehydration. Corundum peaks (cr) are contamination from the alumina shear piston.

rarely be the dominant phase within the metamorphosed crust. However, because lawsonite occurs in veins (Extended Data Fig. 1a shows the lawsonite vein from which we obtained our starting material) or in lawsonite-rich layers<sup>29</sup>, we conclude that dehydration of lawsonite can directly trigger unstable fault slip in the subducting oceanic crust.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 17 August; accepted 26 November 2015.**

- Kita, S., Okada, T., Nakajima, J., Matsuzawa, T. & Hasegawa, A. Existence of a seismic belt in the upper plane of the double seismic zone extending in the along-arc direction at depths of 70–100 km beneath NE Japan. *Geophys. Res. Lett.* **33**, <http://dx.doi.org/10.1029/2006GL028239> (2006).
- Abers, G. A., Nakajima, J., van Keken, P. E., Kita, S. & Hacker, B. R. Thermal-petrological controls on the location of earthquakes within subducting plates. *Earth Planet. Sci. Lett.* **369–370**, 178–187 (2013).
- Chernak, L. J. & Hirth, G. Syndeformational antigorite dehydration produces stable fault slip. *Geology* **39**, 847–850 (2011).
- Hacker, B. R. Subduction factory 2. Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? *J. Geophys. Res.* **108**, <http://dx.doi.org/10.1029/2001JB001129> (2003).
- Marschall, H. R., Altherr, R. & Rüpke, L. Squeezing out the slab—modelling the release of Li, Be and B during progressive high-pressure metamorphism. *Chem. Geol.* **239**, 323–335 (2007).

- Abers, G. A., van Keken, P. E., Kneller, E. A., Ferris, A. & Stachnik, J. C. The thermal structure of subduction zones constrained by seismic imaging: Implications for slab dehydration and wedge flow. *Earth Planet. Sci. Lett.* **241**, 387–397 (2006).
- van Keken, P., Kita, S. & Nakajima, J. Thermal structure and intermediate-depth seismicity in the Tohoku-Hokkaido subduction zones. *Solid Earth* **3**, 355–364 (2012).
- Hacker, B. R.  $H_2O$  subduction beyond arcs. *Geochem. Geophys. Geosyst.* **9**, <http://dx.doi.org/10.1029/2007gc001707> (2008).
- Raleigh, C. & Paterson, M. Experimental deformation of serpentinite and its tectonic implications. *J. Geophys. Res.* **70**, 3965–3985 (1965).
- Gasc, J. *et al.* Simultaneous acoustic emissions monitoring and synchrotron X-ray diffraction at high pressure and temperature: calibration and application to serpentinite dehydration. *Phys. Earth Planet. Inter.* **189**, 121–133 (2011).
- Getsinger, A. J. & Hirth, G. Amphibole fabric formation during diffusion creep and the rheology of shear zones. *Geology* **42**, 535–538 (2014).
- Paterson, M. S. & Olgaard, D. L. Rock deformation tests to large shear strains in torsion. *J. Struct. Geol.* **22**, 1341–1358 (2000).
- Newton, R. C. & Kennedy, G. Some equilibrium reactions in the join  $CaAl_2Si_2O_8-H_2O$ . *J. Geophys. Res.* **68**, 2967–2983 (1963).
- Ulmer, P. & Trommsdorff, V. Serpentine stability to mantle depths and subduction-related magmatism. *Science* **268**, 858–861 (1995).
- Frost, H. J. & Ashby, M. F. *Deformation-Mechanism Maps: the Plasticity and Creep of Metals and Ceramics* 49–51 (Pergamon, 1982).
- Schmidt, M. W. & Poli, S. The stability of lawsonite and zoisite at high pressures: experiments in CASH to 92 kbar and implications for the presence of hydrous phases in subducted lithosphere. *Earth Planet. Sci. Lett.* **124**, 105–118 (1994).
- Jung, H., Fei, Y., Silver, P. G. & Green, H. W. Frictional sliding in serpentine at very high pressure. *Earth Planet. Sci. Lett.* **277**, 273–279 (2009).
- Paterson, M. S. & Wong, T.-f. *Experimental Rock Deformation—The Brittle Field* (Springer, 2005).
- Hadizadeh, J. & Tullis, J. Cataclastic flow and semi-brittle deformation of anorthosite. *J. Struct. Geol.* **14**, 57–63 (1992).
- Schubnel, A. *et al.* Deep-focus earthquake analogs recorded at high pressure and temperature in the laboratory. *Science* **341**, 1377–1380 (2013).
- Hilairet, N. *et al.* High-pressure creep of serpentinite, interseismic deformation, and initiation of subduction. *Science* **318**, 1910–1913 (2007).
- Scholz, C. H. *The Mechanics of Earthquakes and Faulting* 81–90 (Cambridge Univ. Press, 2002).
- Wada, I. & Wang, K. Common depth of slab-mantle decoupling: reconciling diversity and uniformity of subduction zones. *Geochem. Geophys. Geosyst.* **10**, <http://dx.doi.org/10.1029/2009gc002570> (2009).
- Proctor, B. & Hirth, G. Role of pore fluid pressure on transient strength changes and fabric development during serpentine dehydration at mantle conditions: implications for subduction-zone seismicity. *Earth Planet. Sci. Lett.* **421**, 1–12 (2015).
- Hirose, F., Nakajima, J. & Hasegawa, A. Three-dimensional seismic velocity structure and configuration of the Philippine Sea slab in southwestern Japan estimated by double-difference tomography. *J. Geophys. Res.* **113**, <http://dx.doi.org/10.1029/2007JB005274> (2008).
- Kelemen, P. B. & Hirth, G. A periodic shear-heating mechanism for intermediate-depth earthquakes in the mantle. *Nature* **446**, 787–790 (2007).
- Katayama, I., Terada, T., Okazaki, K. & Tanikawa, W. Episodic tremor and slow slip potentially linked to permeability contrasts at the Moho. *Nature Geosci.* **5**, 731–734 (2012).
- Tsuji, Y., Nakajima, J. & Hasegawa, A. Tomographic evidence for hydrated oceanic crust of the Pacific slab beneath northeastern Japan: implications for water transportation in subduction zones. *Geophys. Res. Lett.* **35**, <http://dx.doi.org/10.1029/2008GL034461> (2008).
- Kim, D., Katayama, I., Michibayashi, K. & Tsujimori, T. Rheological contrast between glaucophane and lawsonite in naturally deformed blueschist from Diablo Range, California. *Island Arc* **22**, 63–73 (2013).
- Forneris, J. F. & Holloway, J. R. Phase equilibria in subducting basaltic crust: implications for  $H_2O$  release from the slab. *Earth Planet. Sci. Lett.* **214**, 187–201 (2003).

**Acknowledgements** We thank A. Schubnel, D. Forsyth, T. Altman and T. Togo for technical advice on the acoustic emissions measurement, B. Proctor for discussion and assistance in the laboratory, J. Boesenberg for technical assistance with the electron probe microanalyser, R. Milliken and K. Robertson for helping with XRD measurements, and H. Marschall and Terry E. Tullis for discussions. This study was supported by the US National Science Foundation (EAR-1049582, EAR-1315784).

**Author Contributions** G.H. proposed the project. K.O. conducted experiments and microstructural and mechanical analyses. K.O. took all the photographs. Both authors contributed to developing the main ideas, designing the experiments, interpreting the results and writing the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.O. (keishi\_okazaki@brown.edu).

## METHODS

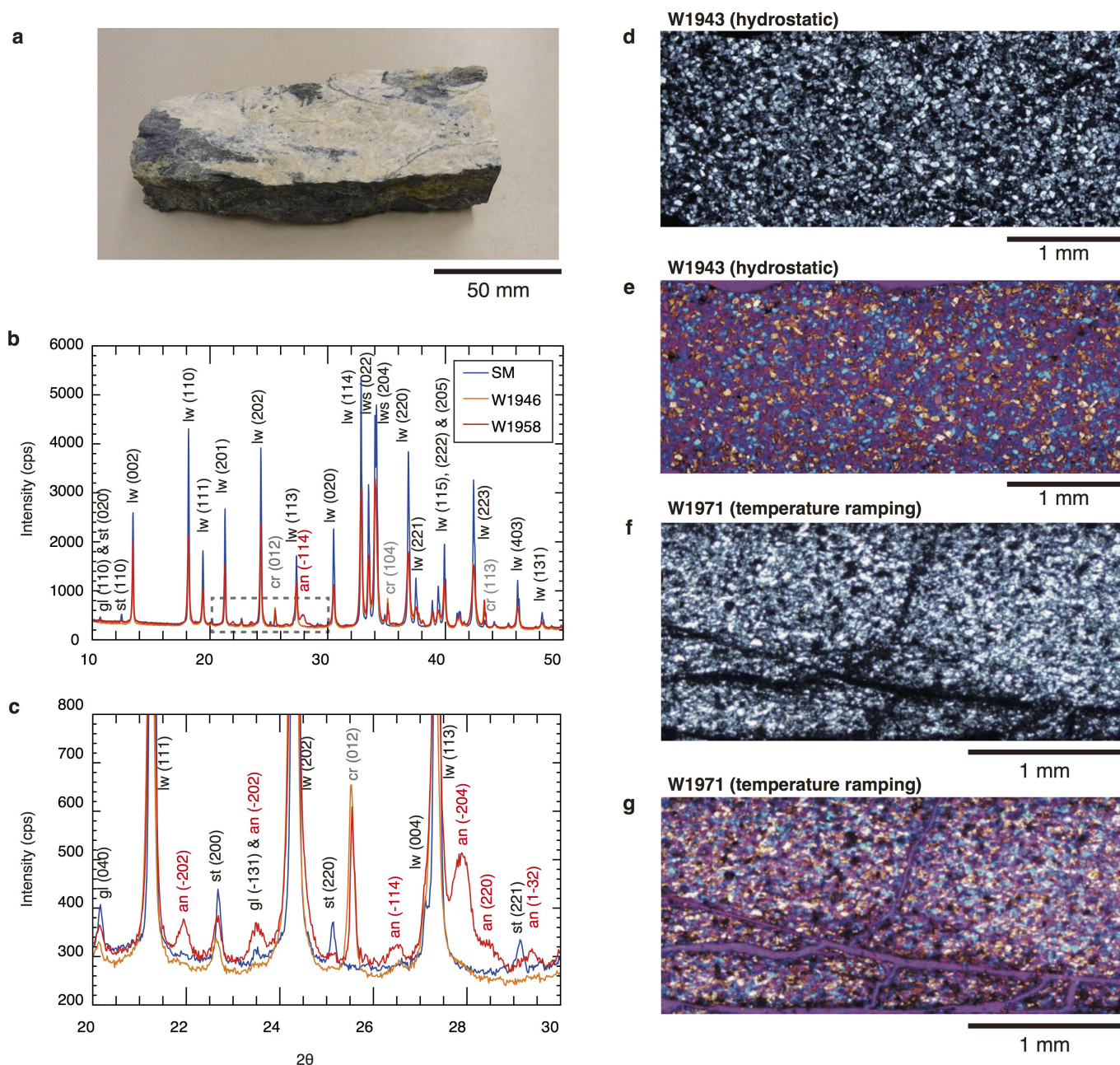
**Starting material and recovered samples.** We used a lawsonite vein within a lawsonite-blueschist block from the Franciscan Complex in California, USA, for the experiments (see photograph in Extended Data Fig. 1a). The lawsonite sample comprises almost pure lawsonite (~98%), with minor staurolite and glaucophane, and the host lawsonite-blueschist block is composed of glaucophane, lawsonite, phengitic mica, and minor titanite. XRD spectra for the lawsonite starting sample show the major lawsonite peaks, and minor staurolite and glaucophane peaks (Fig. 4f and Extended Data Fig. 1). The antigorite samples were derived from a serpentinite block sampled from the Nagasaki metamorphic belt in Japan; this sample is predominantly antigorite (~98%) with minor diopside, spinel and magnetite<sup>31,32</sup>. The lawsonite vein and the antigorite block were crushed and sieved to prepare gouge samples with a grain size of 20  $\mu\text{m}$  to 45  $\mu\text{m}$ . The weight of gouge sample used for each run was 150.0 mg, resulting in a sample thickness of 1.1 mm without any optically visible porosity (<1%) after hot-pressing at the initial experimental conditions. Microstructural observations of recovered samples from the temperature ramping experiments indicate that strain is highly localized along fault surfaces with slickensides, with a lineation parallel to the slip direction. Neither hot-pressed powdered samples nor the recovered samples from the temperature ramping experiments show any notable crystallographic preferred orientation, based on inserting the gypsum plate under cross-polarized light (Extended Data Fig. 1e and g). Similar slickensides are reported from deformation experiments on antigorite within its stability field; all experiments on antigorite at high pressure, including this study, show velocity-strengthening behaviour both within and above the stability field<sup>3,24,33</sup>.

**Experimental conditions.** The experiments were conducted using a Griggs-type apparatus at Brown University. Details of the experimental assembly and

conditions are given in Extended Data Fig. 3 and Extended Data Table 1. Most of the experimental procedures follow protocols used in previous studies<sup>3,24</sup>. Shear deformation was imposed on a thin layer of the sample gouge sandwiched between two alumina or yttria-stabilized zirconia pistons (6.35 mm in diameter) with oblique surfaces oriented 45° from the cylinder's axis. Specimens were mechanically sealed inside a silver jacket with platinum disks placed between the specimen and the pistons. Sodium chloride (NaCl) was used as the confining-pressure medium both inside and outside the graphite furnace. The axial stress  $\sigma_1$  and axial displacement are measured with an external load cell and an external linear variable differential transformer, respectively. The resolution for the stress measurement is  $\pm 25$  MPa and the precision for that is  $\sim 1$  MPa. Load values were corrected for piston friction using methods described in previous studies<sup>24</sup>. Sample temperature was measured using a type S (Pt–Pt10%Rh) thermocouple shielded in mullite placed next to the jacket at the sample centre. Acoustic emissions are measured with a piezoelectric sensor attached to the base plate (see Extended Data Fig. 3a) of the apparatus with a sampling frequency of 2.5 MHz, then high-pass filtered at 100 kHz.

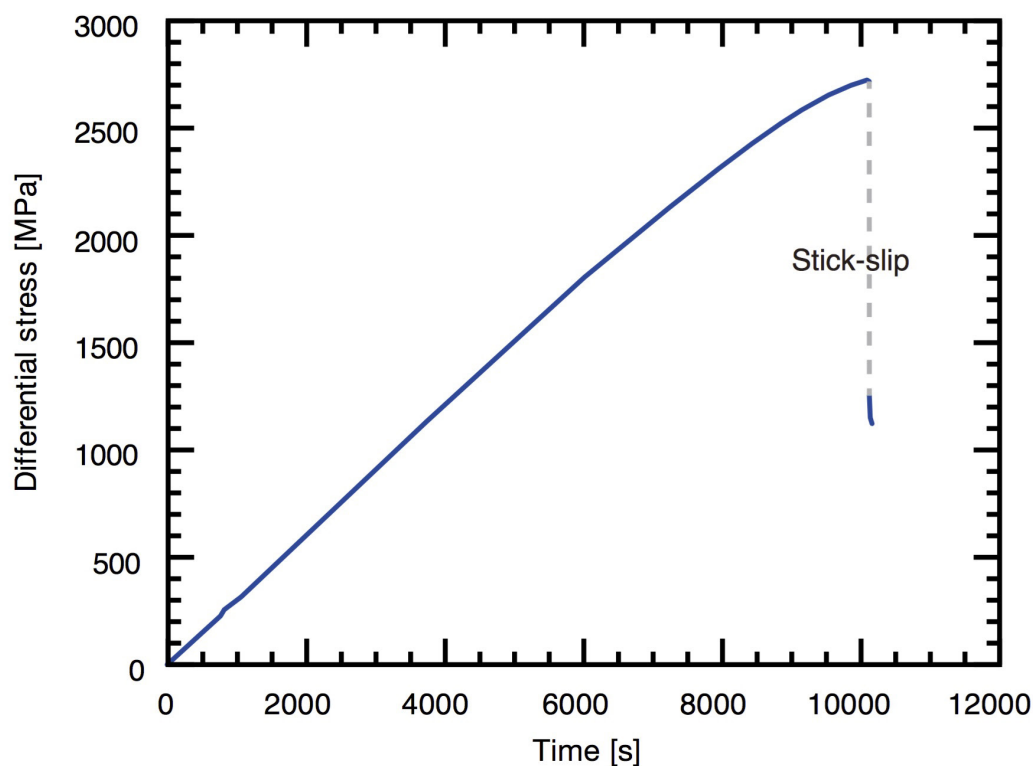
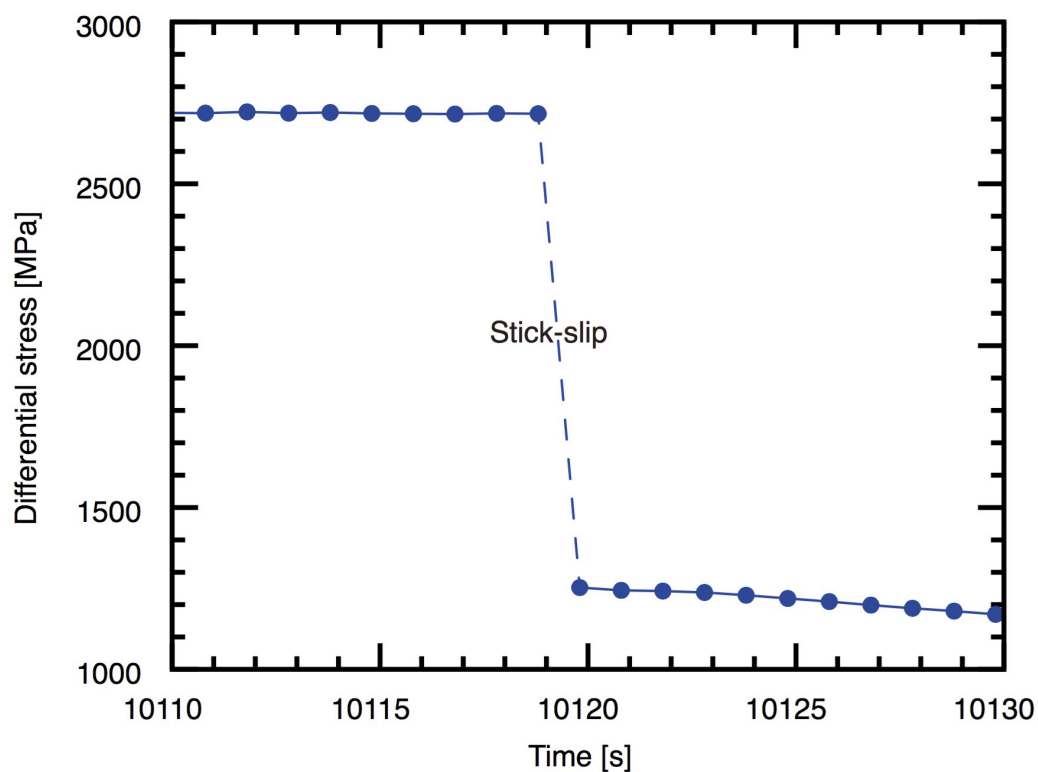
31. Sawai, M., Katayama, I., Hamada, A., Maeda, M. & Nakashima, S. Dehydration kinetics of antigorite using in situ high-temperature infrared microspectroscopy. *Phys. Chem. Miner.* **40**, 319–330 (2013).
32. Okazaki, K. & Katayama, I. Slow stick slip of antigorite serpentinite under hydrothermal conditions as a possible mechanism for slow earthquakes. *Geophys. Res. Lett.* **42**, 1099–1104 (2015).
33. Chernak, L. J. & Hirth, G. Deformation of antigorite serpentinite at high temperature and pressure. *Earth Planet. Sci. Lett.* **296**, 23–33 (2010).
34. Perrillat, J. *et al.* Kinetics of antigorite dehydration: a real-time X-ray diffraction study. *Earth Planet. Sci. Lett.* **236**, 899–913 (2005).





**Extended Data Figure 1 | Lawsonite vein sample and XRD spectra for samples.** **a**, The lawsonite vein in a lawsonite–blueschist block from the Franciscan Complex in California, USA, used for the starting material for the temperature ramp experiments. **b** and **c**, XRD spectra for the starting material (SM) and recovered samples (W1946 and W1965) from experiments on simulated lawsonite gouge. Most of the peaks in the SM are lawsonite (lw); there are minor peaks of staurolite (st) and glaucophane (gl). Spectra for the recovered samples from a temperature ramping experiment (W1965) show a weaker peak in lawsonite than in the SM and

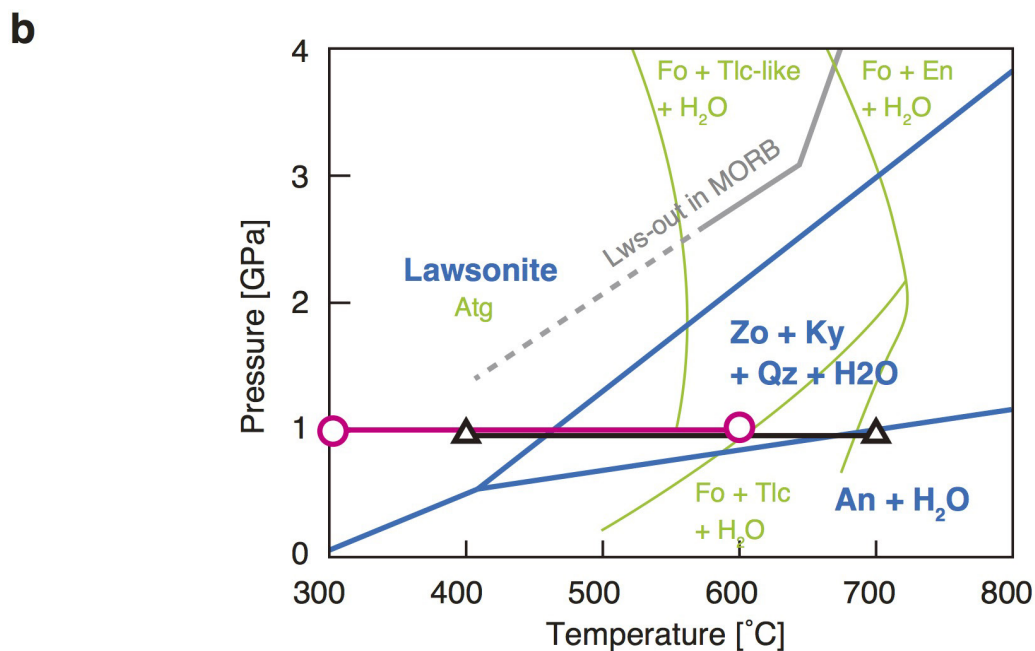
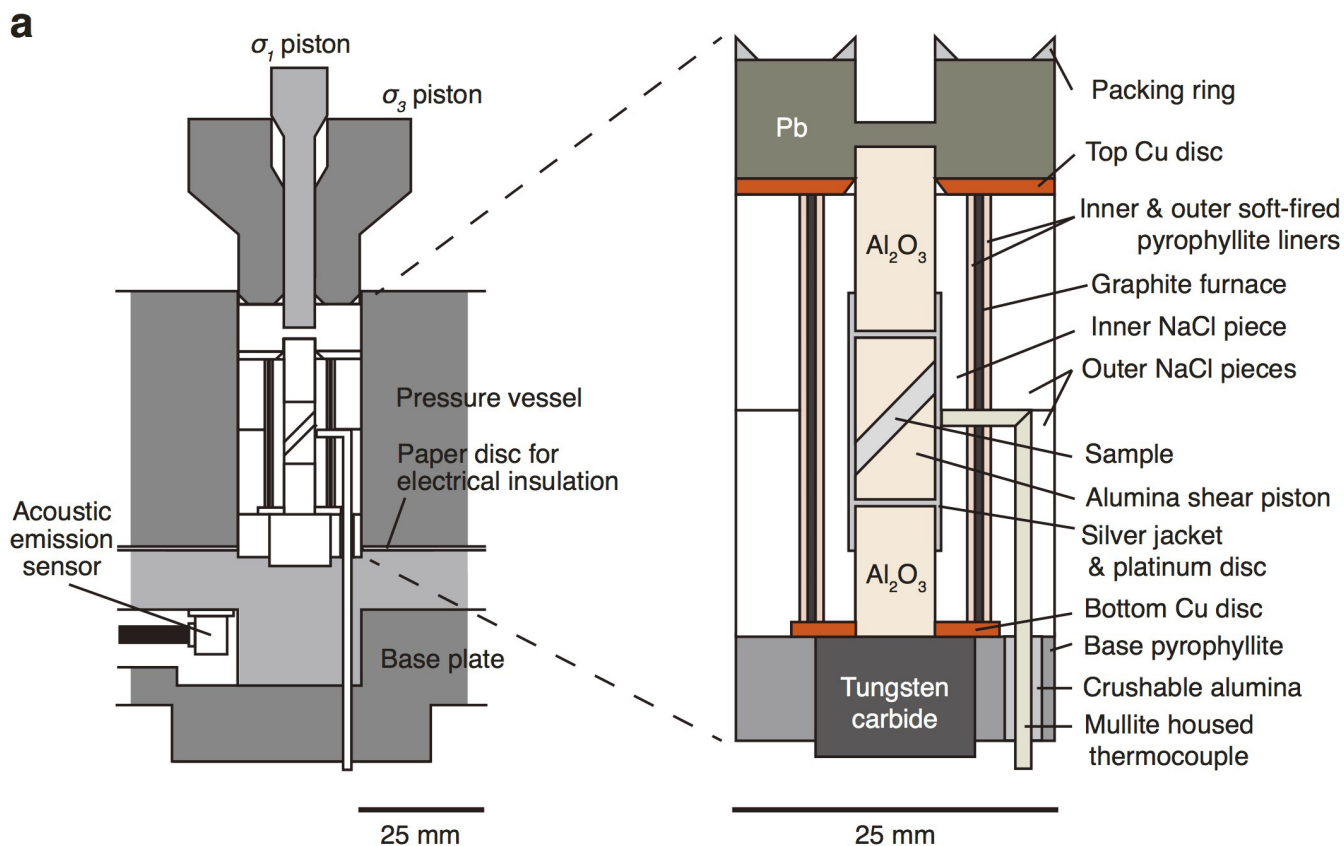
small peaks for anorthite (an). Anorthite, which is formed by lawsonite break down (Fig. 1 and Extended Data Fig. 3), is evidence of lawsonite dehydration. Corundum peaks (cr) are contamination from the alumina shear piston. We did not find any reaction products in the recovered sample deformed within the lawsonite stability field with no temperature ramping (W1946). **d**, **f**, Cross-polarized light micrographs of a hot-pressed sample (W1943) and a sample from a temperature ramping experiment (W1971), respectively. **e**, **g**, Same regions as **d** and **f**, but with the gypsum plate inserted under cross-polarized light. cps, counts per second.

**a****b**

**Extended Data Figure 2 | Stick-slip behaviour of lawsonite gouge without temperature ramping.** **a**, A violent stick-slip of lawsonite was observed in an experiment (W1946) conducted with no temperature ramping at 400 °C, at a confining pressure of 1.0 GPa, and an axial displacement rate of  $0.18 \mu\text{m s}^{-1}$  (Extended Data Table 1). Because of the

enormously large stress drop and fast slip, the thermocouple failed, and the load point displacement was not measured correctly during the stick-slip event, so the unloading slope for this run could not be estimated. **b**, Enlargement of the stick-slip event.





**Extended Data Figure 3 | Sample assembly used in this study (a) and pressure-temperature paths for the temperature ramping experiments (b).** Pink circle and black triangle show pressure-temperature paths for lawsonite and for antigorite, respectively. Blue lines show the lawsonite phase diagram for the  $\text{CaO-SiO}_2\text{-H}_2\text{O}$  system<sup>13</sup> applicable for the experimental system in this study. Grey lines show the lawsonite stability

limit in the MORB system<sup>30</sup> applicable for the natural subduction zone setting. Green lines show the phase diagram for antigorite in the  $\text{MgO-SiO}_2\text{-H}_2\text{O}$  system<sup>14,34</sup>. Atg = antigorite, Zo = zoisite, Ky = kyanite, Qz = quartz, An = anorthite, Fo = forsterite, Tlc = talc, En = enstatite, Tlc-like = talc-like phase.

**Extended Data Table 1 | Experimental details on the temperature ramping experiment on lawsonite**

Run No.	Pressure* [GPa]	Temperature* [°C]	Temperature ramping rate [°C/s]	Sample	Displacement rate [ $\mu\text{m/s}$ ]	Equivalent strain rate [1/s]	(T. ramp rate)/ (strain rate) [°C]	Unloading slope [kN/mm]	Unloading slope [GPa/mm]	Piston Material	AE measurement
W1943	1	300	N/A	Lawsonite	hydrostatic	-	-	-	-	Al <sub>2</sub> O <sub>3</sub>	-
W1945†	1	300	N/A	Lawsonite	0.183	8.6E-05	-	-	-	Al <sub>2</sub> O <sub>3</sub>	-
W1946‡	1	400	N/A	Lawsonite	0.183	8.6E-05	-	-	-	Al <sub>2</sub> O <sub>3</sub>	-
W1949	1.00 (1.13)	300 (600)	0.50	Lawsonite	hydrostatic	-	-	-	-	Al <sub>2</sub> O <sub>3</sub>	-
W1951§	1.00 (1.13)	300 (600)	0.50	Lawsonite	< 0.00183	< 8.6E-7	> 5.8E+5	-263.2 ± 4.0	-8.31 ± 0.13	Al <sub>2</sub> O <sub>3</sub>	-
W1957	1.00 (1.17)	300 (600)	0.50	Lawsonite	0.0183	8.6E-06	5.8E+04	-276.4 ± 4.3	-8.73 ± 0.14	Al <sub>2</sub> O <sub>3</sub> with grooves	-
W1958	0.99 (1.14)	300 (600)	0.50	Lawsonite	0.183	8.6E-05	5.8E+03	-256.3 ± 1.2	-8.09 ± 0.04	Al <sub>2</sub> O <sub>3</sub> with grooves	-
W1962	1.01 (1.19)	300 (600)	0.05	Lawsonite	0.183	8.6E-05	5.8E+02	-266.0 ± 32.9	-8.40 ± 1.04	Al <sub>2</sub> O <sub>3</sub> with grooves	-
W1965	1.01 (1.18)	300 (600)	0.05	Lawsonite	0.0183	8.6E-06	5.8E+03	-249.4 ± 22.6	-7.88 ± 0.71	Al <sub>2</sub> O <sub>3</sub> with grooves	-
W1968	1.03 (1.17)	400 (700)	0.50	Antigorite	0.183	8.6E-05	5.8E+03	-146.4 ± 2.7	-4.62 ± 0.09	Al <sub>2</sub> O <sub>3</sub> with grooves	-
W1971	0.99 (1.11)	300 (600)	0.50	Lawsonite	0.183	8.6E-05	5.8E+03	-257.6 ± 4.1	-8.13 ± 0.13	Al <sub>2</sub> O <sub>3</sub> with grooves	x
W1975	1.05 (1.18)	300 (600)	0.50	NaCl salt	0.183	8.2E-05	6.1E+03	-87.6 ± 8.9	-2.77 ± 0.28	Al <sub>2</sub> O <sub>3</sub> with grooves	x
W1977	1.00 (1.13)	300 (600)	0.50	Antigorite	0.183	8.6E-05	5.8E+03	-154.5 ± 3.4	-4.88 ± 0.11	Al <sub>2</sub> O <sub>3</sub> with grooves	x
W1981	1.00 (1.14)	300 (600)	0.50	Aluminum	0.183	1.4E-05	3.7E+04	-117.1 ± 3.0	-3.70 ± 0.10	Al <sub>2</sub> O <sub>3</sub>	x
W1999¶	1.02 (1.17)	300 (600)	0.50	Lawsonite	0.183	8.6E-05	5.8E+03	-254.5 ± 26.0	-8.04 ± 0.82	ZrO <sub>2</sub> with grooves	x
W2000	1.04 (1.19)	400 (700)	0.50	Antigorite	0.183	8.6E-05	5.8E+03	-192.5 ± 6.2	-6.08 ± 0.20	ZrO <sub>2</sub> with grooves	x

\*Values in brackets show pressures and temperatures after the temperature ramping.

†Deformation was stopped before the yield point,  $\sigma_d \approx 2$  GPa.

§Displacement rate of  $0.18 \mu\text{m s}^{-1}$  for 2 hours prior to the temperature ramping (the axial loading was then stopped during temperature ramping).

We used a displacement rate of  $1.8 \times 10^{-3} \mu\text{m s}^{-1}$  and an equivalent strain rate of  $8.6 \times 10^{-7} \text{s}^{-1}$  for plotting in Fig. 3, which is a maximum value.

||Axial compression configuration.

¶Zirconia piston yielded.

# Historical nectar assessment reveals the fall and rise of floral resources in Britain

Mathilde Baude<sup>1,2†</sup>, William E. Kunin<sup>3</sup>, Nigel D. Boatman<sup>4</sup>, Simon Conyers<sup>4</sup>, Nancy Davies<sup>1,2</sup>, Mark A. K. Gillespie<sup>3†</sup>, R. Daniel Morton<sup>5</sup>, Simon M. Smart<sup>5</sup> & Jane Memmott<sup>1,2</sup>

**There is considerable concern over declines in insect pollinator communities and potential impacts on the pollination of crops and wildflowers<sup>1–4</sup>. Among the multiple pressures facing pollinators<sup>2–4</sup>, decreasing floral resources due to habitat loss and degradation has been suggested as a key contributing factor<sup>2–8</sup>. However, a lack of quantitative data has hampered testing for historical changes in floral resources. Here we show that overall floral rewards can be estimated at a national scale by combining vegetation surveys and direct nectar measurements. We find evidence for substantial losses in nectar resources in England and Wales between the 1930s and 1970s; however, total nectar provision in Great Britain as a whole had stabilized by 1978, and increased from 1998 to 2007. These findings concur with trends in pollinator diversity, which declined in the mid-twentieth century<sup>9</sup> but stabilized more recently<sup>10</sup>. The diversity of nectar sources declined from 1978 to 1990 and thereafter in some habitats, with four plant species accounting for over 50% of national nectar provision in 2007. Calcareous grassland, broadleaved woodland and neutral grassland are the habitats that produce the greatest amount of nectar per unit area from the most diverse sources, whereas arable land is the poorest with respect to amount of nectar per unit area and diversity of nectar sources. Although agri-environment schemes add resources to arable landscapes, their national contribution is low. Owing to their large area, improved grasslands could add substantially to national nectar provision if they were managed to increase floral resource provision. This national-scale assessment of floral resource provision affords new insights into the links between plant and pollinator declines, and offers considerable opportunities for conservation.**

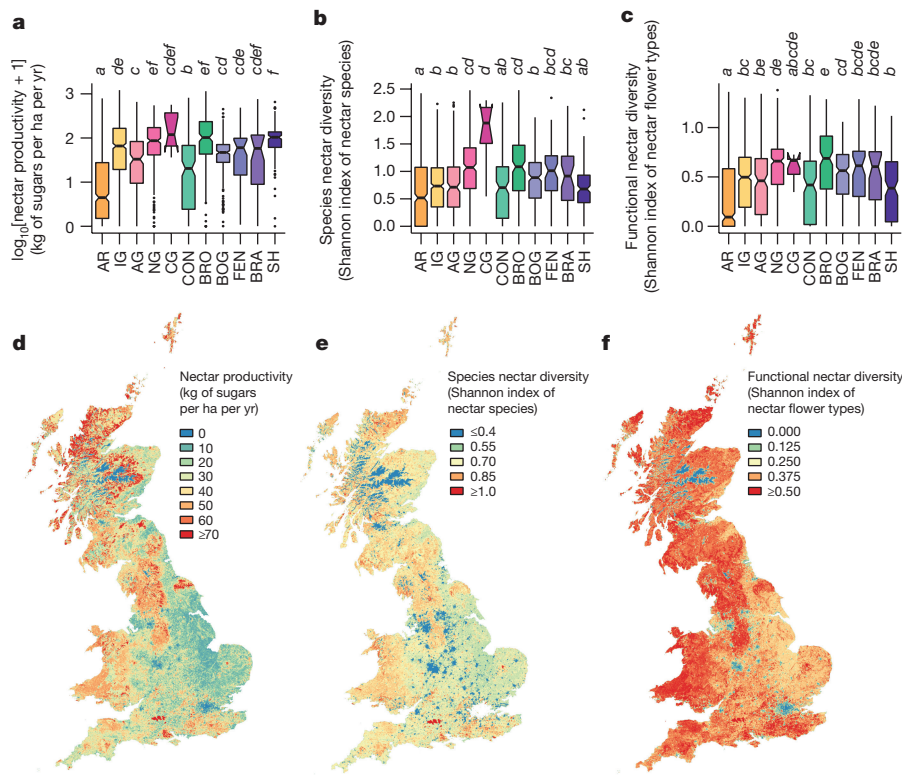
Concerns have been raised about declines in both wild and managed insect pollinators<sup>1–4</sup>. Although several potential drivers have been cited<sup>2–4</sup>, one important factor in pollinator declines may be the loss of floral resources due to changes in land use and management<sup>5–8</sup>. Several factors may have caused decreased floral resources in Great Britain and other developed countries, including increased use of herbicides<sup>11</sup>, destruction of traditional landscape features such as hedgerows<sup>12</sup> and loss and degradation of wildflower-rich natural habitats<sup>13–15</sup>. Current strategies to mitigate pollinator declines focus primarily on enhancing floral resources<sup>4</sup>, including agri-environmental scheme options such as sowing nectar flower mixtures<sup>16,17</sup>. There is evidence for declines in some key pollinator forage plants in Great Britain<sup>5</sup> and The Netherlands<sup>7</sup>, but the notion that the overall availability of floral resources has declined is largely based on subjective assessments. Floral resources have never been quantified at national or even landscape scales.

While both nectar and pollen are important floral resources, we focus on nectar because of its importance as an energy source in the diets of adult bees, and because it provides a common currency (total sugars) in which we can express the nutritional contribution of all plant

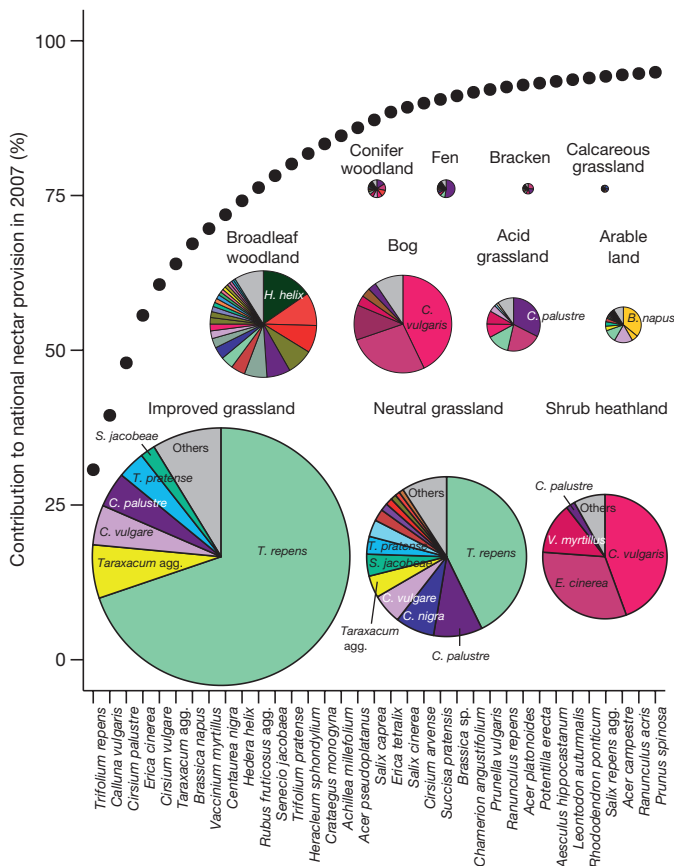
species<sup>18</sup>. We quantified the nectar resources in Great Britain by combining directly measured and modelled nectar productivity data per unit cover for 260 common plant species (Supplementary Table 11) with historical vegetative cover estimates from the British Countryside Survey<sup>19</sup>, a representative national-scale survey of plant community composition. Together, the 260 species comprise the vast majority of British nectar sources as they include virtually all nectar-producing plants from the set of species covering 99% of the British land area. Using vegetation data from the latest Countryside Survey (2007), we quantified recent nectar productivity of habitats (nectar sugar per unit area and time) and the diversity of their nectar sources (considering nectar production both by species and by floral morphology groups, referred to as ‘species nectar diversity’ and ‘functional nectar diversity’, respectively). Production was scaled up to estimate national nectar provision using the estimated area of habitats<sup>19</sup>, allowing the contributions of species, habitats and agri-environment schemes to national nectar provision to be assessed. We estimated historical shifts in nectar provision over recent decades using data from earlier Countryside Survey rounds (1978, 1990, 1998 and 2007), considering both changes in nectar productivity within habitats and changes in habitat area. We also investigated floral resource changes from the 1930s onward for England and Wales, based solely on changes in habitat coverage.

Considering the most recent Countryside Survey (2007), there are significant differences in annual nectar productivity, species nectar diversity and functional nectar diversity among habitats (Extended Data Table 1). Calcareous grassland, broadleaved woodland and neutral grassland are the best in all three respects (as well as shrub heathland for nectar productivity only), whereas arable land is consistently the poorest habitat (Supplementary Table 1). These habitat differences in nectar value create geographical variation in nectar productivity and diversity across Great Britain (Fig. 1). After taking into account the national land cover of habitats, improved grassland contributed most (29%) to potential national nectar supply in 2007. Four species of plant, *Trifolium repens*, *Calluna vulgaris*, *Cirsium palustre* and *Erica cinerea*, together produce over 50% of nectar nationally (see Extended Data Table 2 and Supplementary Information for further information about these species and their pollinators), and 22 species produce over 90% (Fig. 2). Other species may of course be important for pollen provision. A consideration of flowering phenology reveals seasonal variation nationally (Fig. 3): 60% of nectar is provided in July and August when the flower density of British dominant species peaks. Because heathland species are unlikely to contribute as much to nectar provision in other European countries, this seasonal pattern may differ. The relative nectar value of linear features (hedgerows, watersides and road verges) depends on habitat. With the exception of those in shrub heathland and bog, linear features produce more nectar per unit area compared with nonlinear features (the contrast is particularly high in landscapes dominated by arable land, improved grassland and conifer woodland;

<sup>1</sup>School of Biological Sciences, University of Bristol, Life Sciences Building, Bristol BS8 1TQ, UK. <sup>2</sup>Cabot Institute, University of Bristol, Bristol BS8 1UJ, UK. <sup>3</sup>School of Biology, University of Leeds, Leeds LS2 9JT, UK. <sup>4</sup>Fera Science Ltd., Sand Hutton, York YO41 1LZ, UK. <sup>5</sup>NERC Center for Ecology & Hydrology, Bailrigg, Lancaster LA1 4AP, UK. <sup>†</sup>Present address: Collegium Sciences et Techniques EA 1207 LBLGC, Université d'Orléans, F-45067, Orléans, France (M.B.); Department of Science and Engineering, Sogn og Fjordane University College, 6851 Sogndal, Norway (M.A.K.G.).



**Figure 1 | Nectar productivity and diversity in Great Britain in 2007.** **a**, Box plots of  $\log_{10}(x+1)$  nectar productivity (kg of sugars per ha per year) per habitat. **b**, Box plots of species nectar diversity (Shannon index of nectar species) per habitat. **c**, Box plots of functional nectar diversity (Shannon index of nectar flower types) per habitat. Box plots are based on 2007 nonlinear vegetation data (see Supplementary Table 1 for sample sizes). Habitat types (AR, arable land; IG, improved grassland; AG, acid grassland; NG, neutral grassland; CG, calcareous grassland; CON, conifer woodland; BRO, broadleaf woodland; BOG, bog; FEN, fen; BRA, bracken; SH, shrub heathland) significantly different from one another are indicated by different italicised letters (Tukey multiple comparisons tests). See Extended Data Table 1 for ANOVA results. See 'Statistical analyses' section of the Methods for detailed statistical methods and definition of box plot elements. **d**, Map of nectar productivity. **e**, Map of species nectar diversity. **f**, Map of functional nectar diversity. Maps are based on 2007 land cover and nonlinear vegetation data. Habitat nectar values for mapping correspond to statistical estimates from linear mixed effects models (Supplementary Table 1).

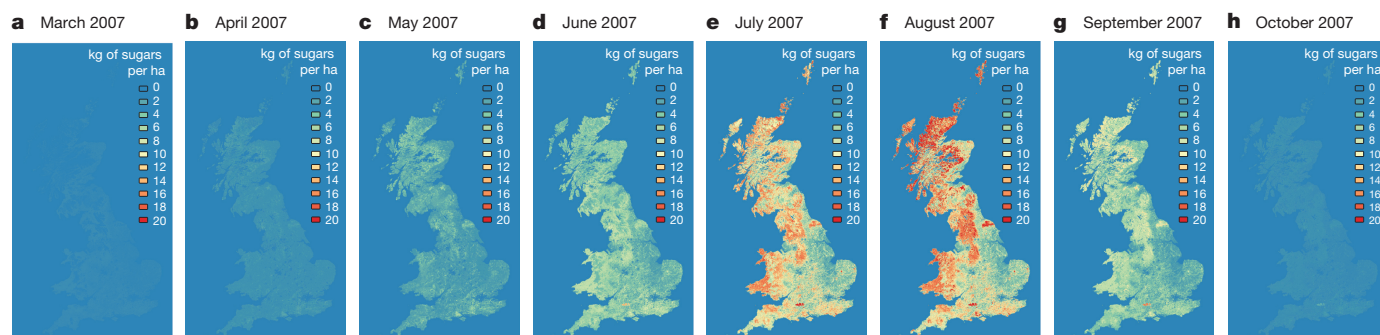


**Figure 2 | Plant species' contributions to Great Britain nectar provision and to habitat nectar provision, based on 2007 land cover and nonlinear vegetation data.** The dotted line represents the cumulative contribution of plant species to the national nectar provision in 2007 (only species that contribute to the first 95% are shown). The pie charts represent the contribution of plant species towards nectar production in each habitat (only the species that contribute to the first 90% are shown) in 2007. The size of each pie chart is proportional to the contribution of each habitat to national nectar provision in 2007.

Extended Data Fig. 1). Of the five types of agri-environment scheme options we investigated, nectar flower mixtures have the highest nectar productivity value, followed by enhanced margins (Extended Data Table 3). Nectar flower mixture options are similar to hedgerows in terms of annual nectar productivity per unit area, but they cover a much smaller area, and consequently contribute far less to the national nectar resources (0.1% of nectar supply comes from nectar flower mixtures compared to 3% from hedgerows in England; Extended Data Table 3).

Historical shifts in nectar productivity, species nectar diversity and functional nectar diversity over recent decades depended on the habitat type and time period considered (Extended Data Table 1). From 1978 to 1990, annual nectar productivity decreased significantly in arable land and conifer woodland, but from 1990 to 1998, none of the habitats showed significant changes in nectar productivity. From 1998 to 2007, nectar productivity increased significantly in arable land and neutral grassland (Extended Data Fig. 2). Nectar diversity, both at the level of plant species and functional groups, decreased significantly in arable land and improved grassland from 1978 to 2007. Species nectar diversity also significantly decreased in conifer woodland and broadleaved woodland during that period. From 1978 to 1990, species nectar diversity declined in all habitats (except bog), significantly so in arable land and conifer woodland; thereafter it remained roughly constant, except in arable land where it increased significantly from 1998 to 2007 (see Extended Data Fig. 2 and Supplementary Information for details on functional nectar diversity). For the 1930s we have information only on shifts in land cover (but not floral abundances within them), and only for England and Wales<sup>20</sup>. Assuming no change in floral composition within habitats, we found a strong decline in national nectar provision from the 1930s to 1978 (−32%), followed by a period of stagnation from 1978 to 2007 (Fig. 4 and Supplementary Table 2). Incorporating shifts in nectar productivity within habitats for recent decades showed an increase in national nectar provision from 1998 to 2007 (+51% in England and Wales and +25% for Great Britain as a whole; Fig. 4 and Supplementary Table 3). While shifts in vegetation composition within dominant habitats predominate as causes of recent increases, no quantitative data are available before 1978. This recent upturn could be caused by decreased acidification<sup>21</sup>, decreased nitrogen deposition<sup>22</sup> and agricultural set-asides<sup>23</sup> during this period (Supplementary Table 4).

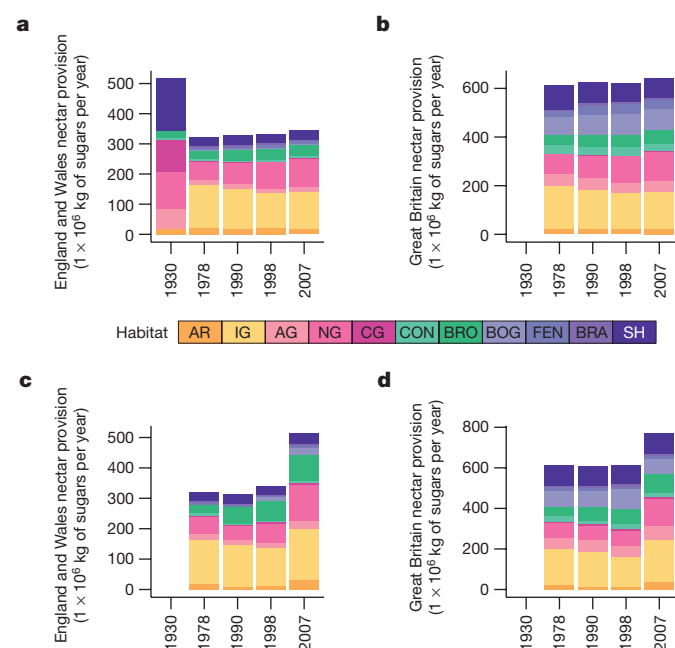




**Figure 3 | Seasonal nectar productivity in Great Britain, based on 2007 land cover and nonlinear vegetation data.** a–h, Maps of nectar productivity in kg of sugars per ha from March to October 2007. Hot colours correspond to high nectar productivity while cold colours correspond to low nectar productivity (see colours scale). Note that urban areas are assigned with nectar productivity values equal to zero, hence the

blue colours in cities. Nectar productivity values for mapping correspond to back-transformed estimates of the linear mixed model fitted on  $\log_{10}(x+1)$  nectar productivity of 2007 Countryside Survey nonlinear plots with habitat, month and their interaction as fixed effects and plots nested within squares as random effects.

However, post-war changes in habitat management (for example, herbicide use in arable land, cessation of woodland coppicing, nitrogen deposition in grasslands; Supplementary Table 4) almost certainly resulted in lower nectar per unit area, suggesting that our estimates of losses based on land use change alone are conservative; actual resource declines may have been much larger than the recent increases (see Supplementary Information). Owing to their large area, improved grassland provided the greatest contribution to the increase in national nectar provision from 1998 to 2007 (Extended Data Fig. 3). After discounting the contribution of *Trifolium repens* in improved grasslands, as it may not flower in heavily grazed fields, the increase in nectar provision from 1998 to 2007 remained (Supplementary Information and Extended Data Fig. 4).



**Figure 4 | Historical changes in nectar provision (in  $1 \times 10^6$  kg of sugars per year) at the national scale in England and Wales (1930–2007) and in Great Britain (1978–2007).** a, b, Nectar provision partitioned by habitat, based on land cover for 1930 (England and Wales only), 1978, 1990, 1998 and 2007, using vegetation data from 1978 for all years (assuming unchanged nectar productivity within habitats across time) in England and Wales (a) and Great Britain (b). c, d, Nectar provision partitioned by habitat, based on land cover and vegetation data for 1978, 1990, 1998 and 2007 in England and Wales (c) and Great Britain (d). See Fig. 1 for habitat type codes and Supplementary Table 5 for habitat land cover values.

The historical pattern of change in nectar resources closely parallels documented shifts in pollinator communities (Extended Data Fig. 5). Substantial declines in floral resources and their diversity in the mid to late twentieth century, when agricultural intensification peaked, coincide with a period of heightened pollinator extinctions<sup>9</sup>. The stabilization and partial recovery of resources in recent decades corresponds to concomitant periods of decelerated declines and partial recovery in some pollinator groups<sup>10</sup>.

Our findings provide new evidence based on floral resources to support habitat conservation and restoration. First, we provide evidence of the high nectar value of calcareous grassland for pollinating insects. Calcareous grassland area has declined drastically in Great Britain, and only a small fraction of the historical national cover remained by 2007 (refs 13, 14). Second, the low availability and diversity of nectar sources in arable habitats highlights the need to provide supplementary resources to support pollination services in farmlands, especially as the use of insect-pollinated crops has increased nationally<sup>24</sup> and globally<sup>25</sup>. The conservation and restoration of broadleaf woodland and neutral grassland as components of the farmland matrix could help to support diverse flower-visiting insect communities in arable land. The contrast in nectar productivity between linear features and the surrounding vegetation is particularly high in arable land, suggesting that linear features, especially hedgerows, provide an efficient means to enhance floral resources in farmlands if they are managed appropriately to allow flowering<sup>26</sup>. While agri-environment options such as nectar flower mixtures can also enhance the supply of floral resources locally, their contribution to nectar provision nationally remains low. The higher profile given to floral resource provision in the revised Countryside Stewardship guidelines for England<sup>16</sup> may substantially enhance resources in future. Finally, our results indicate that improved grassland has the potential to contribute massively to the nectar available nationally. Small adjustments to the management cycle in improved grasslands, allowing white clover, the dominant resource species, to flower, would help realize this potential, although its utility might be restricted to a limited number of pollinator species (Extended Data Table 2). Together, our results on the nectar values of the commonest British plants and the historical changes in plant communities provide the evidence base needed to understand recent national changes in nectar provision and identify the management options needed to restore national nectar supplies.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 6 June; accepted 11 December 2015.**

1. Biesmeijer, J. *et al.* Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* **313**, 351–354 (2006).
2. Potts, S. G. *et al.* Global pollinator declines: trends, impacts and drivers. *Trends Ecol. Evol.* **25**, 345–353 (2010).
3. Vanbergen, A. J. & the Insect Pollinators Initiative. Threats to an ecosystem service: pressures on pollinators. *Front. Ecol. Environ.* **11**, 251–259 (2013).
4. Goulson, D., Nicholls, E., Botías, C. & Rotheray, E. L. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* **347**, 1255957 (2015).
5. Carvell, C. *et al.* Declines in forage availability for bumblebees at a national scale. *Biol. Conserv.* **132**, 481–489 (2006).
6. Roulston, T. H. & Goodell, K. The role of resources and risks in regulating wild bee populations. *Annu. Rev. Entomol.* **56**, 293–312 (2011).
7. Scheper, J. *et al.* Museum specimens reveal loss of pollen host plants as key factor driving wild bee decline in The Netherlands. *Proc. Natl Acad. Sci. USA* **111**, 17552–17557 (2014).
8. Kleijn, D. & Raemakers, I. A retrospective analysis of pollen host plant use by stable and declining bumble bee species. *Ecology* **89**, 1811–1823 (2008).
9. Ollerton, J., Erenler, H., Edwards, M. & Crockett, R. Extinctions of aculeate pollinators in Britain and the role of large-scale agricultural changes. *Science* **346**, 1360–1362 (2014).
10. Carvalheiro, L. G. *et al.* Species richness declines and biotic homogenisation have slowed down for NW-European pollinators and plants. *Ecol. Lett.* **16**, 870–878 (2013).
11. Robinson, R. A. & Sutherland, W. J. Post-war changes in arable farming and biodiversity in Great Britain. *J. Appl. Ecol.* **39**, 157–176 (2002).
12. Petit, S., Stuart, R. C., Gillespie, M. K. & Barr, C. J. Field boundaries in Great Britain: stock and change between 1984, 1990 and 1998. *J. Environ. Manage.* **67**, 229–238 (2003).
13. Blackstock, T. H. *et al.* The extent of semi-natural grassland communities in lowland England and Wales: a review of conservation surveys 1978–96. *Grass Forage Sci.* **54**, 1–18 (1999).
14. Ratcliffe, D. A. Post-medieval and recent changes in British vegetation: the culmination of human influence. *New Phytol.* **98**, 73–100 (1984).
15. Fuller, R. M. The changing extent and conservation interest of lowland grasslands in England and Wales: A review of grassland surveys 1930–84. *Biol. Conserv.* **40**, 281–300 (1987).
16. Natural England. *Countryside Stewardship Manual* <https://www.gov.uk/guidance/countryside-stewardship-manual> (2015).
17. Carvell, C., Meek, W. R. W., Pywell, R. F., Goulson, D. & Nowakowski, M. Comparing the efficiency of agri-environment schemes to enhance bumble bee abundance and diversity on arable field margins. *J. Appl. Ecol.* **44**, 29–40 (2007).
18. Willmer, P. *Pollination and Floral Ecology* 1–778 (Princeton Univ. Press, 2011).
19. Carey, P. D. *et al.* Countryside Survey: UK Results from 2007. 1–105 (NERC/Centre for Ecology & Hydrology, 2008).
20. Stamp, L. D. *The Land of Britain: its Use and Misuse* (Longmans, Green and Co., 1948).
21. Kirk, G. J. D., Bellamy, P. H. & Lark, R. M. Changes in soil pH across England and Wales in response to decreased acid deposition. *Glob. Change Biol.* **16**, 3111–3119 (2006).
22. Reynolds, B. *et al.* Countryside Survey: National “Soil Change” 1978–2007 for Topsoils in Great Britain — Acidity, Carbon, and Total Nitrogen Status. *Vadose Zone J.* **12**, (2013).
23. Boatman, N. D., Jones, N. E., Conyers, S. T. & Pietravalle, S. Development of plant communities on set-aside in England. *Agric. Ecosyst. Environ.* **143**, 8–19 (2011).
24. Breeze, T. D., Bailey, P., Balcombe, K. G. & Potts, S. G. Pollination services in the UK: How important are honeybees? *Agric. Ecosyst. Environ.* **142**, 137–143 (2011).
25. Aizen, M. A., Garibaldi, L. A., Cunningham, S. A. & Klein, A. M. Long-term global trends in crop yield and production reveal no current pollination shortage but increasing pollinator dependency. *Curr. Biol.* **18**, 1572–1575 (2008).
26. Staley, J. T. *et al.* Changes in hedgerow floral diversity over 70 years in an English rural landscape, and the impacts of management. *Biol. Conserv.* **167**, 97–105 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This research was supported by the UK Insect Pollinators Initiative (IPI) ‘AgriLand: Linking agriculture and land use change to pollinator populations’ project, funded by the Biotechnology and Biological Sciences Research Council (BBSRC), Wellcome Trust, Scottish Government, Department of Environment, Food and Rural Affairs (DEFRA) and Natural Environment Research Council (NERC) under the auspices of the Living with Environmental Change partnership: grant BB/H014934/1 (<http://www.agriland.leeds.ac.uk>). Land Cover and Countryside Survey data are owned by NERC – Centre for Ecology & Hydrology (<http://www.countrysidesurvey.org.uk>).

**Author Contributions** The study was conceived by W.E.K. and J.M. The field survey was carried out by M.B. and N.D. with the help of J.M. The data were compiled and analysed by M.B. with suggestions from W.E.K., J.M., S.M.S., R.D.M. and M.A.K.G. Vegetation data from the Countryside Survey database were extracted by S.M.S. Agri-environment scheme data were provided and analysed by N.D.B. and S.C. The national maps were generated by R.D.M. All authors discussed the results and contributed during manuscript writing.

**Author Information** The floral resource database will be made available from the NERC Environmental Information Data Centre (<http://dx.doi.org/10.5285/69402002-1676-4de9-a04e-d17e827db93c> and <http://dx.doi.org/10.5285/6c6d3844-e95a-4f84-a12e-65be4731e934>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. ([mathilde.baude@univ-orleans.fr](mailto:mathilde.baude@univ-orleans.fr)).

## METHODS

**Constructing the nectar database by scaling up nectar resources from the flower to the vegetative scale.** *Identifying the key plant species to be sampled.* Although there are >2,800 plant species in Great Britain<sup>27</sup>, only 1,341 of them are common enough to have been encountered in the Countryside Survey. Of these, the 454 commonest species accounted for 99% of national plant cover in 2007. More than half of these 454 species are unrewarding to pollinators (mainly bryophytes, pteridophytes, gymnosperms and wind-pollinated angiosperms<sup>28</sup>), leaving 220 species that are likely to contribute substantially to floral resources at a national scale. We focus here on these 220 species, along with an additional 50 species that we believe to be locally important floral sources (for example, *Buddleja davidii*, *Impatiens glandulifera*, *Knaulia arvensis*). Together, these 270 plant species provide a focal set of potential importance in national nectar provision (Supplementary Table 11). *Quantifying nectar productivity empirically: the 'surveyed species'.* Of the 270 species, 175 were surveyed in the field from February 2011 to October 2012, mainly in the south of England. When possible (112 species), nectar was collected from plants in at least two populations in two locations. For three species (*Caltha palustris*, *Lamium purpureum* and *Sinapis arvensis*), half the nectar samples, and for *Viola arvensis* all the samples, were collected from pot-grown plants, because insufficient flowering field populations were found. For the remaining species, nectar was collected from plants in one field population. When possible, the different populations were sampled on different dates, thus providing some measure of variation due to differences in location and weather. Note that nectar was collected in only 1–2 sites per species, and so intraspecific variation in production per flower was not assessed (but see Supplementary Information).

Nectar was collected from ten single flowers in each population between 09:00 and 16:00 h (median 20 and range 5–30 flowers collected per species in total; see Extended Data Fig. 6 and Supplementary Information for site correlation); these had been bagged (using  $1.4 \times 1.7$  mm fabric mesh) for 24 h to prevent depletion by nectar-feeding insects. When possible (76 species), glass microcapillaries (1 and 5  $\mu$ l Minicaps, Hirshmann, Eberstadt, Germany) were used directly to collect the nectar, otherwise single flowers were rinsed twice with 1–5  $\mu$ l of distilled water added to the nectaries with a pipette for 1 min, and the diluted nectar solution was collected. The sugar concentration of nectar (%; g sucrose per 100 g solution) was measured by using a hand-held refractometer modified for small volumes (Eclipse, Bellingham and Stanley, Tunbridge Wells, UK). The amount of sugar produced per flower basis over 24 h (s;  $\mu$ g of sugars per flower per 24 h) was calculated using the formula<sup>29</sup>  $s = 10dvC$  where  $v$  is the volume collected ( $\mu$ l), and  $d$  is the density of a sucrose solution at a concentration  $C$  (g sucrose per 100 g solution) as read on the refractometer. The density of the sucrose solution was calculated by the formula<sup>29</sup>  $d = 0.0037921C + 0.0000178C^2 + 0.9988603$ .

The number of open flowers per unit area of vegetative cover (flower density) was estimated for 179 species by placing five quadrats (0.5 m  $\times$  0.5 m) haphazardly on each flowering population (median 10 quadrats, range 1–20 quadrats; see Extended Data Fig. 6 and Supplementary Information for site correlation). In each quadrat, we counted the number of open floral units of the focal species (a 'floral unit' is one or multiple flowers that can be visited by insects without flying<sup>30</sup>; for example a composite flower head of daisy, *Bellis perennis*). We also counted the number of open flowers present in one typical open floral unit in each quadrat. Vegetative cover for each plant species was estimated using a point-quadrat approach with the cross-strings of the quadrat: cover was expressed as proportional to the number of the 36 cross-strings covered by the foliage of the species of interest in each quadrat. For trees, instead of using quadrats, we counted the number of floral units in a 3D cube (0.5  $\times$  0.5  $\times$  0.5 m) that was placed in the outer areas of foliage. This was extrapolated to the whole column situated above the unit of vegetative cover by measuring the height of tree foliage with an inclinometer (PM-5/360 PC Suunto) and by estimating the distribution of the flowers within the tree foliage (subjectively assessed scores: from 1 for a strongly biased flower distribution on the outer edges of the foliage to 5 for a homogeneous full flower distribution). Given that flower density is not constant throughout the flowering season, we estimated variations in flower density according to a triangular function from the estimated peak of flowering through the flowering season which was documented from recorded phenologies<sup>28,31,32</sup> (see Supplementary Information and Extended Data Fig. 6 for phenology parameter relationships). An alternative nectar rectangular phenology productivity database was also generated by keeping nectar productivity of each species constant throughout the flowering season; this was used to perform sensitivity analyses.

The mean nectar sugar content from a single flower (produced over a 24 h period) was multiplied up to the nectar content of a single floral unit (number of flowers in a floral unit), then to the amount of nectar per unit area (number of flowers per m<sup>2</sup>), to the amount of nectar per unit area for each month (variation in flower density over the flowering season) and finally to the amount of nectar per unit area per year. Despite relatively low sample sizes per species compared to

species-specific studies, our estimates of sugar production were well correlated with published values both per flower per day and per area per year (Extended Data Fig. 6 and Supplementary Information). This empirical method provided the nectar productivity values for 161 plant species among the 175 initially surveyed (nectar productivity could not be scaled up for some species due to mismatches with phenological data, see Supplementary Information).

*Modelling nectar productivity: the 'unsurveyed species'.* To model the nectar productivity of the plant species that could not be surveyed in the field, we used a predictive modelling approach. We first analysed variation in the nectar values from the surveyed species. A linear model was fitted to annual nectar sugar productivity ( $\log_{10}(x+1)$  transformed) as a function of plant traits. Plants traits were mainly collected from the BioFlor database<sup>33</sup>, and included: 'flower shape', 'breeding system', 'life span', the degree of 'decline', the maximum 'height', the 'flowering period' and 'family' (see Supplementary Information for definitions). The estimates from the most parsimonious statistical model based on AIC criterion (Supplementary Table 6,  $N = 153$ ; adjusted  $r^2 = 0.55$ ) were used to predict the annual nectar sugar productivity for the initial list of surveyed and unsurveyed species on the basis of their traits. To check the validity of the predicted values, we adopted a repeated 'leave-one-out' approach to model successively all the excluded values from the empirically derived data sets. Then, we applied a standardized major axis regression on the  $\log_{10}(x+1)$  transformed empirically derived and modelled nectar values of the surveyed species (Extended Data Fig. 6). We predicted the nectar values for 252 species; and giving priority to empirical and default values, we included 94 of them in our database. An alternative nectar productivity database was also generated by considering only the species with empirical nectar values; this was used to perform sensitivity testing.

*Ascribing default values for nectar productivity.* For four crop species harvested before flowering—onion (*Allium cepa*), cabbage (*Brassica oleracea* cultivated), turnip (*Brassica rapa*) and radish (*Raphanus sativus*)—we assigned a value of zero for nectar productivity. A zero value was also assigned to *Helianthemum nummularium*, despite the missing flower density data, given that we collected no nectar in flowers. In the Countryside Survey vegetation data set, some taxa are only identified at the genus level; we interpreted these taxa to represent the commonest species in the genus (for example, *Centaurea* sp. was interpreted as *Centaurea nigra*). For 10 species out of the initial list of 270 it was not possible to quantify nectar production, leading to a total of 260 species with quantified annual and monthly nectar productivity values (161 values from empirical research, 94 modelled values, and 5 default values, Supplementary Table 11). All the above steps of scaling-up process are summarized in Supplementary Table 7.

**Using the Countryside Survey vegetation database to scale up nectar resources from plant species to communities at the habitat and national scales.** Spatio-temporal variations in nectar provision at the national scale were calculated by combining our nectar productivity data set with vegetation and land cover data already recorded during the Countryside Survey<sup>19</sup>. The Countryside Survey is a national survey of plant communities conducted in 1978, 1990, 1998 and 2007 in Great Britain (England, Wales and Scotland). The survey was conducted by selecting 1-km sample squares at random from 32 land classes<sup>19</sup> representing physiographically similar sampling domains throughout Great Britain, ensuring an unbiased representation of the British non-urban landscape. Within each square, a random, stratified sample of five areal (nonlinear) square plots (200 m<sup>2</sup>) was established and the presence and the percentage cover of all vascular plant species were recorded. These plots were classified to 17 habitat classes, but we only used data from 11 habitats: acid grassland, arable land, bog, bracken, broadleaf woodland, calcareous grassland, conifer, fen, improved grassland, neutral grassland and shrub heath (Supplementary Table 8 for habitat description). The habitats not used were inland rock, littoral rock/supralittoral rock, littoral sediment/supralittoral sediment, montane and urban habitats; these were excluded due to low sample sizes. Even though urban habitats probably contribute to the national nectar provision, we were unable to include this habitat in this study because the Countryside Survey was not designed to survey urban areas. In 1.14% of Countryside Survey plots, two or more habitats were attributed to the same plot; these were excluded for this study. Additional plots were used to sample linear features in each 1 km square, covering hedgerows, streamsides and road verges (1  $\times$  10 m and oriented along the linear feature). Each linear plot was also attributed to its nearest adjacent habitat.

To investigate the most recent nectar patterns, we used the most comprehensive vegetation data set from the Countryside Survey 2007 that encompasses all nonlinear plots (2,576 plots in 2007). To focus on linear features, we included vegetation data from linear features plots (1,951 plots in 2007). To test for historical changes from 1978 to 2007, we used vegetation data from nonlinear plots shared between the 1978, 1990, 1998 and 2007 Countryside Surveys (529 shared plots in England and Wales and 768 in Great Britain; Supplementary Table 9). We focused on the shared plots across years because the Countryside Survey sampling design was



modified over time (for example, from fixed to proportional plot number per land class from 1978 to 1990).

The annual nectar productivity within each plot (kg per ha per year) is the sum of the nectar productivity of each species (kg per ha cover per year) weighted by their vegetative cover in the plot (%), assuming that the vegetative cover is representative of floral abundance (see Extended Data Fig. 7 and Supplementary Information for details). Nectar productivity values of plots were used to statistically estimate the annual nectar productivity for each habitat (kg per ha per year). The annual nectar provision of each habitat (kg per year) was computed from their annual habitat nectar productivity (kg per ha per year) multiplied by their respective national land covers for each survey (areas of habitats in ha from Countryside Surveys<sup>19,34,35</sup>; Supplementary Table 5). These were summed to estimate the annual national nectar provision in 1978, 1990, 1998 and 2007. For the 1930s period, areas of habitats (only available for England and Wales) were derived from the digitalized Dudley Stamp land utilization survey maps<sup>20</sup>; see Supplementary Information and Supplementary Table 5). Because nectar productivity can't be assessed for this period, we quantified nectar provision in 1930, 1978, 1990, 1998 and 2007 assuming unchanged nectar productivity within habitats but using observed shifts in land cover among habitats across time. The national nectar provision of hedgerows was calculated from their mean nectar productivity (kg per ha per year) multiplied by their estimated area in England (length of hedgerows from Countryside Survey 2007 for England<sup>35</sup>, assuming a 1 m width).

The contribution of habitat or species to the national nectar provision in 2007 is the fraction of nectar provided by these entities (in %). The amount of nectar offered by each habitat in 2007 is calculated from habitat nectar productivity (estimated value of habitat productivity) multiplied by its national area. The amount of nectar offered by each species in 2007 is calculated from the sum of its average nectar productivity stratified by habitat and multiplied by habitat national area. The contribution of habitat or species to the historical changes in national nectar provision is expressed by the absolute change (in kg of sugars), which is the difference in the amount of nectar produced by the entity during the time period considered. Relative change (in %) which is the absolute change multiplied by 100 and divided by the amount of nectar produced at the initial date, refers to the magnitude of change for each entity.

Nectar diversity was estimated through two Shannon indexes (using 'vegan' package in R<sup>36</sup>) that encompass both the richness and the evenness of nectar producing sources (see Supplementary Information). The species nectar diversity index, based on the proportion of nectar produced by each species, was calculated as follows:

$$H_{sp}' = - \sum_{i=1}^S p_i \times \ln(p_i)$$

where  $p_i$  is the proportional nectar contribution of plant species  $i$  and  $S$  is the total number of plant species in each plot.

The functional nectar diversity index, based on the proportion of nectar produced by each floral morphology group, reflects the diversity of nectar sources in terms of resource accessibility for flower-visiting insects. Flower types were derived from Müller flower classification system recorded from the BioFlor database<sup>33</sup>, which was condensed into five classes: pollen rewarding flowers, open, partly hidden, hidden, and bee flowers (see Supplementary Information). The functional nectar diversity index was computed as follows:

$$H_{fun}' = - \sum_{i=1}^S p_i \times \ln(p_i)$$

where  $p_i$  is the proportional nectar contribution of flower type  $i$  and  $S$  is the total number of flower types in each plot.

The annual nectar productivity (kg of sugars per ha per year), species nectar diversity (Shannon index of nectar contribution of plant species) and functional nectar diversity (Shannon index of nectar contribution of floral morphology groups) in 2007 were mapped at the British national scale using the Great Britain Land Cover Maps of 2007<sup>37</sup>.

**Using agri-environment scheme flower abundance data to estimate nectar provision within agri-environment scheme options at the national scale.** Various options are available for managing habitats to provide floral resources for pollinators, some of which are eligible for grant aid under European Union funded agri-environment schemes. Agri-environment options within the English 'Environmental Stewardship' scheme included sowing nectar flower mixtures (EF4/HF4), sowing wild bird seed mixtures (EF2/HF2), creation or enhancement of floristically enhanced buffer strips (HE10), re-introduction or continuation of haymaking (haymaking supplement HK18) and creation, restoration and maintenance of species-rich semi-natural grassland (HK6/7/8). These five options were selected as the most likely to provide floral resources for pollinators.

Field study sites were located on farmland and nature reserves in which the following replicates of the pollinator habitats were present: nectar flower mixtures ( $n = 32$ ), wild bird seed mixtures ( $n = 4$ ), enhanced field margins/road verges ( $n = 7$ ), hay meadows ( $n = 5$ ) and species-rich grasslands ( $n = 7$ ). These were existing habitats representing ongoing management by the land owners or land managers concerned. Transects 100 m long  $\times$  6 m wide were established in each habitat. The number of floral units of each flowering species was recorded on 1 to 3 occasions, in  $20 \times 1 \text{ m}^2$  quadrats per transect. Annual nectar productivity (kg of sugars per ha per year) was calculated for each species at each site from the average estimated nectar productivity at the peak of the flowering season derived from the several counts of floral units across the flowering period (analogous to Supplementary Information). The values for the species present in each habitat were then summed to estimate productivity for each habitat.

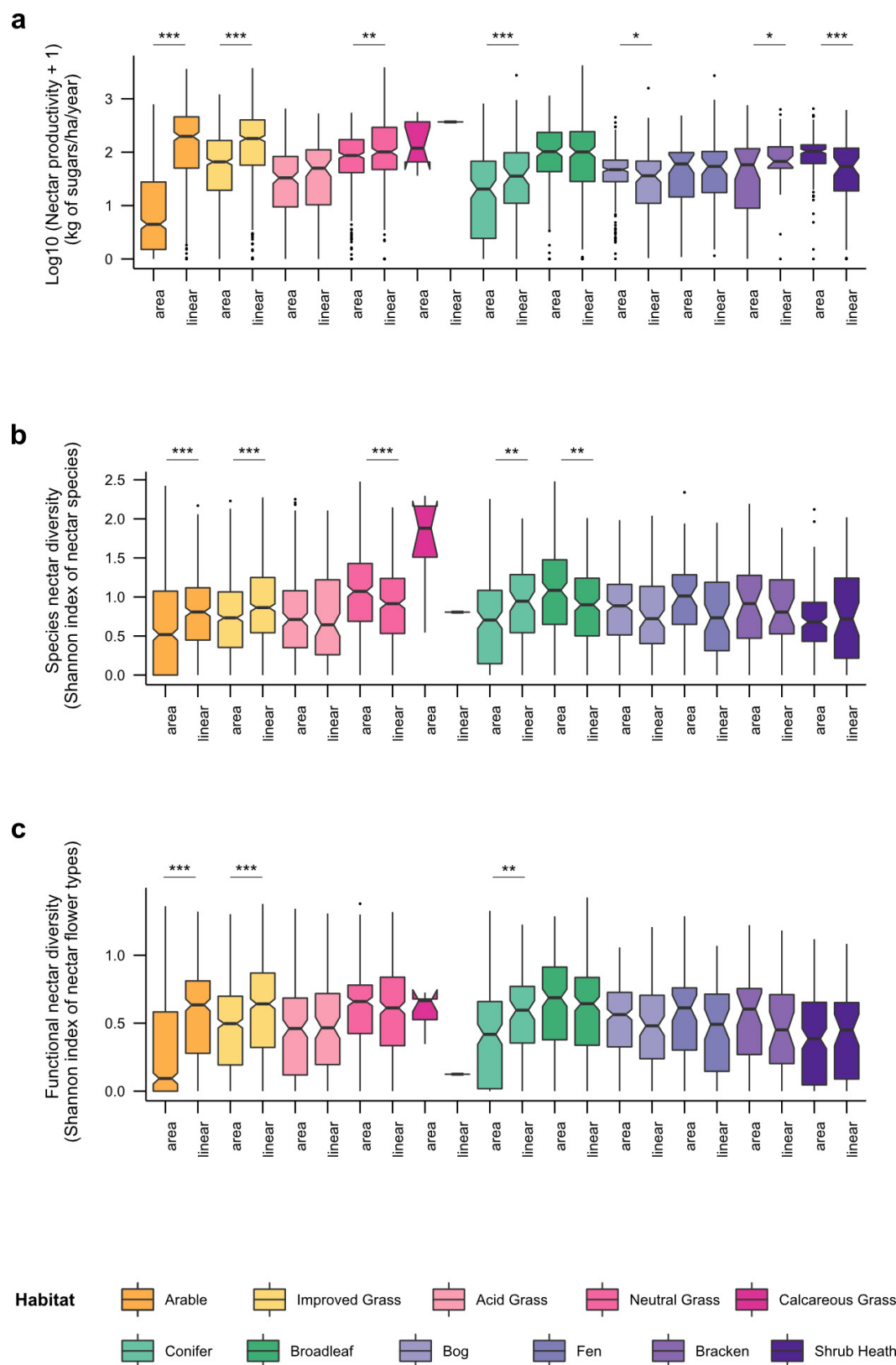
National areas of options providing floral resources in the English agri-environment scheme 'Environmental Stewardship' were extracted for 2007 for England (data for Great Britain were unavailable) from data supplied by Natural England<sup>38,39</sup>. Mean nectar productivity per unit area was multiplied by the national area of each option to give nectar provision by that option (kg of sugars per year). The total contribution of nectar provision provided by Environmental Stewardship in England is a minimum value, as it has been compared to national provision estimated from vegetative cover rather than direct flower counts and we did not take into account the more limited floral resources potentially provided by other options.

**Statistical analyses.** No statistical methods were used to predetermine sample size. Statistical analyses were carried out with Linear Mixed-Effect Models (lme function from 'nlme' package) in R 3.0.1 (ref. 36). To investigate the most recent nectar variations (2007), we analysed the  $\log_{10}(x+1)$  annual nectar productivity, species nectar diversity and functional nectar diversity according to the type of habitat ("HABITAT"; 11 habitats) of the nonlinear plots. The differences in  $\log_{10}(x+1)$  nectar productivity, species nectar diversity and functional nectar diversity between nonlinear and linear features were analysed according to the type of habitat ("HABITAT"; 11 habitats), the type of vegetation surveyed ("TYPE"; nonlinear vs linear features) and the interaction between these two terms. Countryside Survey square ("SQUARE") was included as a random term in these models in order to account for the spatial auto-correlation of plots nested into 1 km squares. In order to investigate historical changes over recent decades (1978–2007), we analysed the  $\log_{10}(x+1)$  annual nectar productivity, species nectar diversity and functional nectar diversity computed from the shared nonlinear plots in 1978, 1990, 1998 and 2007 according to the type of habitat ("HABITAT"), the year ("YEAR") considered as a categorical factor, and the interaction between these two terms. We included plots nested within square ("SQUARE/PLOTS") as random terms to account for the spatial and temporal auto-correlation of the data in this latter model. This latter statistical test was repeated considering all shared plots in Great Britain or only those in England and Wales to provide estimates of habitat nectar productivity across time for distinct areas, allowing comparisons with earlier (1930s) habitat information only available for that latter area. Significant differences among modalities were analysed with multiple comparisons (single-step method adjusted  $P$ -values from glht function in 'multcomp' package in R<sup>36</sup>). Letter-based representation of all multiple comparisons was achieved from multcompLetters function in 'multcompView' package in R<sup>36</sup>. Model residuals were plotted to visually check that normality and homoscedasticity assumptions were satisfied. We re-ran the same analyses with the Countryside Survey vegetation data combined with (1) the alternative nectar rectangular phenology productivity database (created by keeping constant nectar productivity of each species during the flowering season); and (2) using only the empirical nectar productivity database, as sensitivity tests (Extended Data Fig. 4 and Supplementary Information). Plots were performed with ggplot2 package in R<sup>36</sup>. All box plots show the median, 25th and 75th percentiles (lower and upper hinges), trimmed ranges that extend from the hinges to the lowest and highest values within  $1.5 \times$  inter-quartile range of the hinge (lower and upper whiskers) plus outliers (filled circles). Notches that extend  $1.58 \times$  inter-quartile range/square root of the number of observations were represented to give a roughly 95 interval for comparing medians.

27. Preston, C. D., Pearman, D. A. & Dines, T. D. *New Atlas of the British and Irish Flora: An Atlas of the Vascular Plants of Britain, Ireland, The Isle of Man and the Channel Island* 910 (Oxford Univ. Press, 2002).
28. Fitter, A. H. & Peat, H. J. The Ecological Flora Database. *J. Ecol.* **82**, 415–425 (1994).
29. Corbet, S. A. *et al.* Native or exotic? Double or single? Evaluating plants for pollinator-friendly gardens. *Ann. Bot.* **87**, 219 (2001).
30. Carvalheiro, L. G., Barbosa, E. R. M. & Memmott, J. Pollinator networks, alien species and the conservation of rare plants: *Trinia glauca* as a case study. *J. Appl. Ecol.* **45**, 1419–1427 (2008).

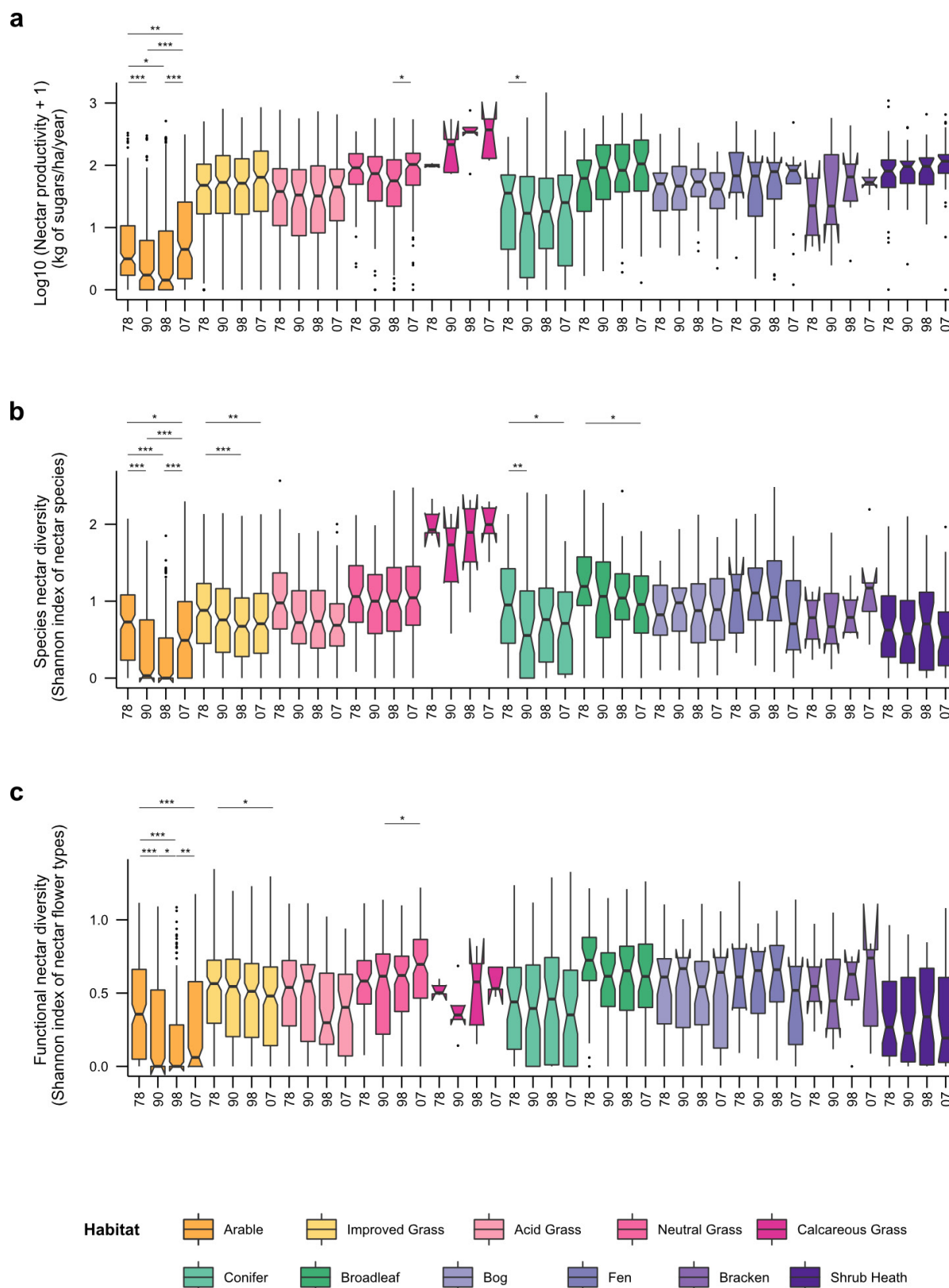


31. Botanical Society of the British Isles. <http://www.botanicalkeys.co.uk/flora/> (2011).
32. Kirk, W. D. J. & Howes, F. N. *Plants for Bees: a Guide to the Plants that Benefit the Bees of the British Isles* 1–311 (International Bee Res. Assoc., 2012).
33. Klotz, S., Kühn, I. & Durka, W. BIOLFLOR - Eine Datenbank zu biologisch-ökologischen Merkmalen der Gefäßpflanzen in Deutschland. <http://www2.ufz.de/biolflor/index.jsp> (2002).
34. Wood, C. M., Howard, D. C., Henrys, P. A. & Smart, S. M. *Countryside Survey: Measuring Habitat Change over 30 years 1978 Data Rescue - Final Report*, 1–18 (2012).
35. *Countryside Survey: England Results from 2007*. NERC/Centre for Ecology & Hydrology, Department for Environment, Food and Rural Affairs, Natural England, 1–119 (2009).
36. R Development Core Team. R: A Language and Environment for Statistical Computing. <http://www.R-project.org> (2013).
37. Morton, R. D. *et al.* Final Report for LCM2007 - the new UK land cover map. Countryside Survey Technical Report No 11/07 NERC/Centre for Ecology & Hydrology (CEH Project Number: C03259), 112 (2011).
38. Department for Environment Food and Rural Affairs. *Entry Level Stewardship Handbook Terms and conditions and how to apply PB10355*. 116 (2005).
39. Department for Environment Food and Rural Affairs. *Higher Level Stewardship Handbook Terms and conditions and how to apply PB10382*. 123 (2005).
40. Raine, N. R. & Chittka, L. Nectar production rates of 75 bumblebee-visited flower species in a German flora (Hymenoptera: Apidae: *Bombus terrestris*). *Entomol. Gen.* **30**, 191–192 (2007).



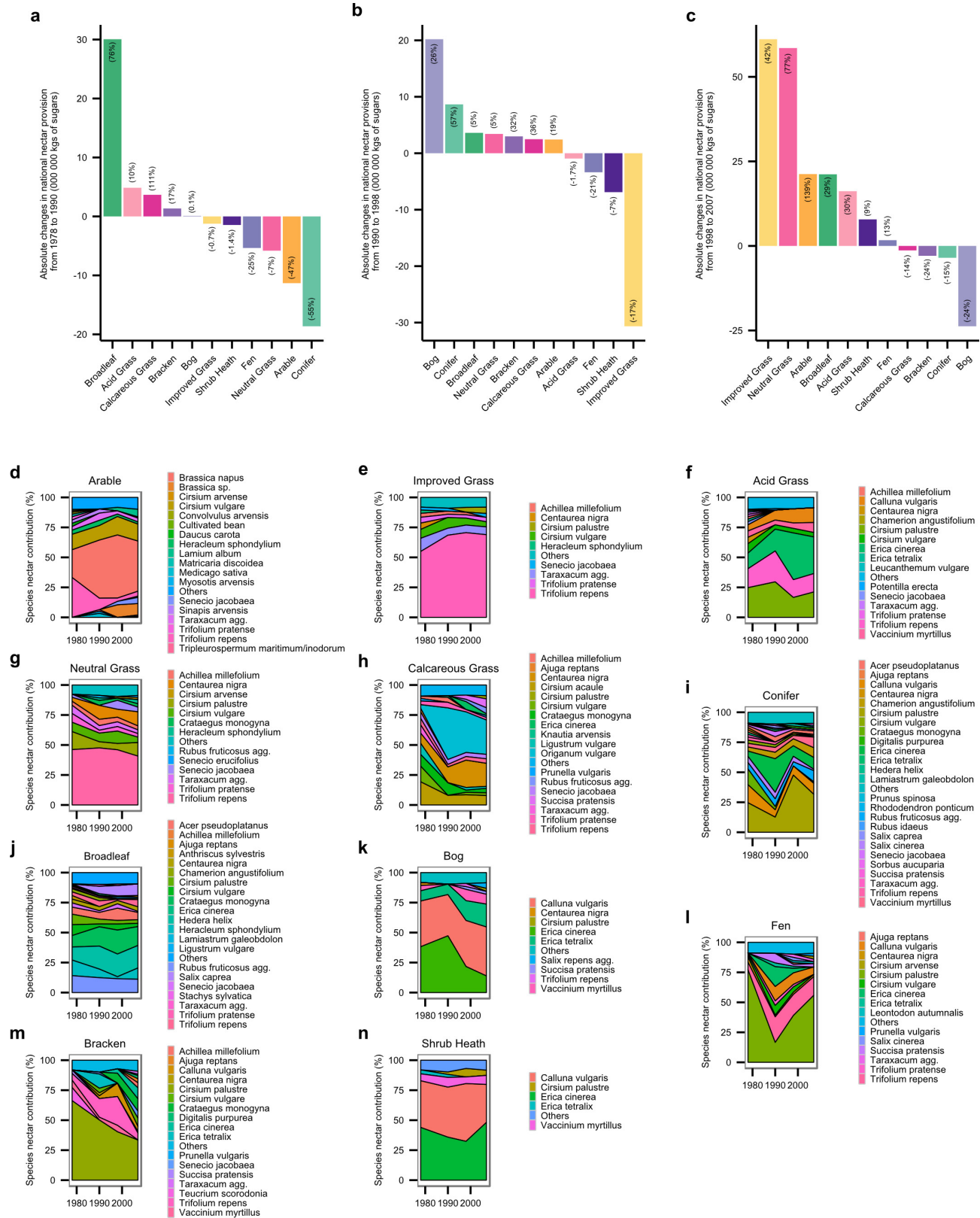
**Extended Data Figure 1 | Annual nectar productivity and diversity in linear features in 2007.** **a**, Box plots of  $\log_{10}(x + 1)$  nectar productivity according to the location of the vegetation surveyed (area (nonlinear) versus linear features) in each habitat. **b**, Box plots of species nectar diversity according to the location of the vegetation surveyed (area versus linear features) in each habitat. **c**, Box plots of functional nectar diversity according to the location of the vegetation surveyed (area versus linear features) in each habitat. Significant differences of locations (area

versus linear features) in habitats are indicated by asterisks as follows:  $*P \leq 0.05$ ;  $**P \leq 0.01$ ;  $***P \leq 0.001$ . Statistical models were re-run without calcareous grassland habitat (to meet residuals homoscedasticity constraint) in order to check that significant effects remained. See Extended Data Table 1 for ANOVA results. See 'Statistical analyses' section of the Methods for detailed statistical methods and definition of box plot elements.



**Extended Data Figure 2 | Historical changes in nectar productivity and diversity per habitat over recent decades (1978 to 2007).** a, Box plots of  $\log_{10}(x + 1)$  nectar productivity per habitat, based on vegetation data for 1978, 1990, 1998 and 2007. b, Box plots of species nectar diversity per habitat, based on vegetation data for 1978, 1990, 1998 and 2007. c, Box plots of functional nectar diversity per habitat, based on vegetation data

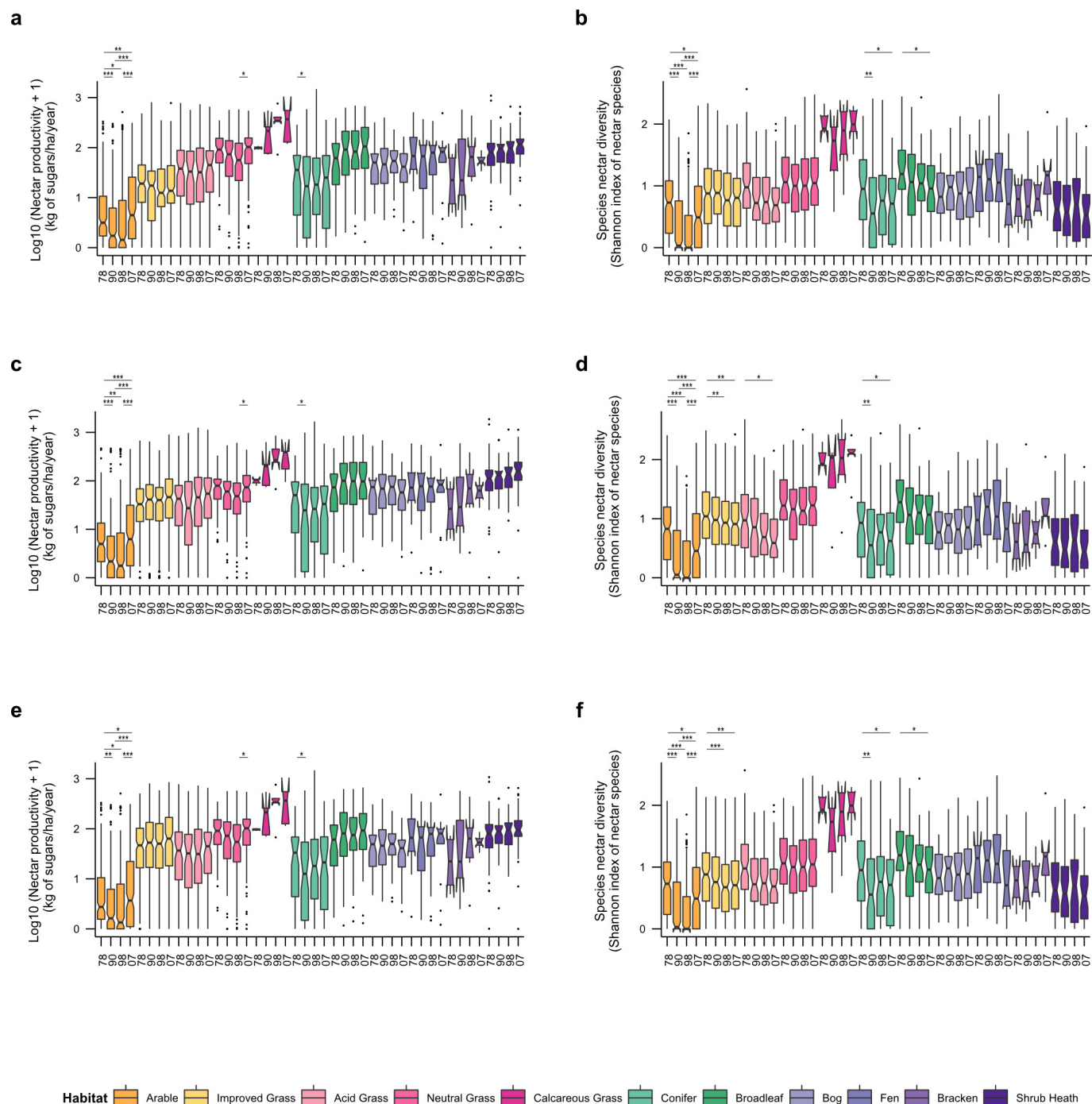
for 1978, 1990, 1998 and 2007. Significant differences of time periods per habitats are indicated by asterisks ( $*P \leq 0.05$ ;  $**P \leq 0.01$ ;  $***P \leq 0.001$ ). See Extended Data Table 1 for ANOVA results and Supplementary Table 3 for sample sizes. See 'Statistical analyses' section of the Methods for detailed statistical methods and definition of box plot elements.



**Extended Data Figure 3 | Habitat contributions to the national nectar provision shifts and species contributions to habitats over recent decades (1978 to 2007).** **a–c**, Habitat contributions to the national nectar provision changes from **a**, 1978 to 1990, **b**, 1990 to 1998, and **c**, 1998 to 2007. All bar plots represent the absolute changes (in  $1 \times 10^6$  kg of sugars)

for each habitat during the time period considered. Numbers in brackets indicate the relative changes (in %). **d–n**, Species contributions to nectar provision in 1978, 1990, 1998 and 2007 per habitat type. Only species that contribute to the first 90% are shown. See Supplementary Table 10 for main contributing species to the national changes from 1978 to 2007.

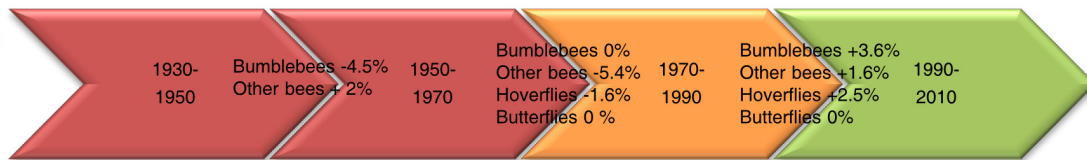




**Extended Data Figure 4 | Sensitivity analyses of historical trends from 1978 to 2007 in nectar productivity and species diversity with alternative data sets.** **a, b**, Box plots of  $\log_{10}(x + 1)$  nectar productivity (**a**) and box plots of species nectar diversity per habitat (**b**) based on vegetation data for 1978, 1990, 1998 and 2007 discounting the contribution of grazed white clover in improved grassland. **c, d**, Box plots of  $\log_{10}(x + 1)$  nectar productivity (**c**) and box plots of species nectar diversity per habitat (**d**), based on vegetation data for 1978, 1990, 1998 and 2007 and computed with the alternative rectangular phenology function.

**e, f**, Box plots of  $\log_{10}(x + 1)$  nectar productivity (**e**) and box plots of species nectar diversity per habitat (**f**), based on vegetation data for 1978, 1990, 1998 and 2007 and computed considering only the species with empirical nectar values. Significant differences of time periods per habitat are indicated by asterisks (\* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ ). See Supplementary Table 3 for sample sizes and Supplementary Information for details. See 'Statistical analyses' section of the Methods for detailed statistical methods and definition of box plot elements.

Carvalho et al. 2014  
What? Changes in species richness at the national scale  
Where? Great Britain  
Database? UK Biological Records Centre



Ollerton et al. 2014  
What? Extinction rates of bees and flower-visiting wasps  
Where? Britain  
Database? BWARS



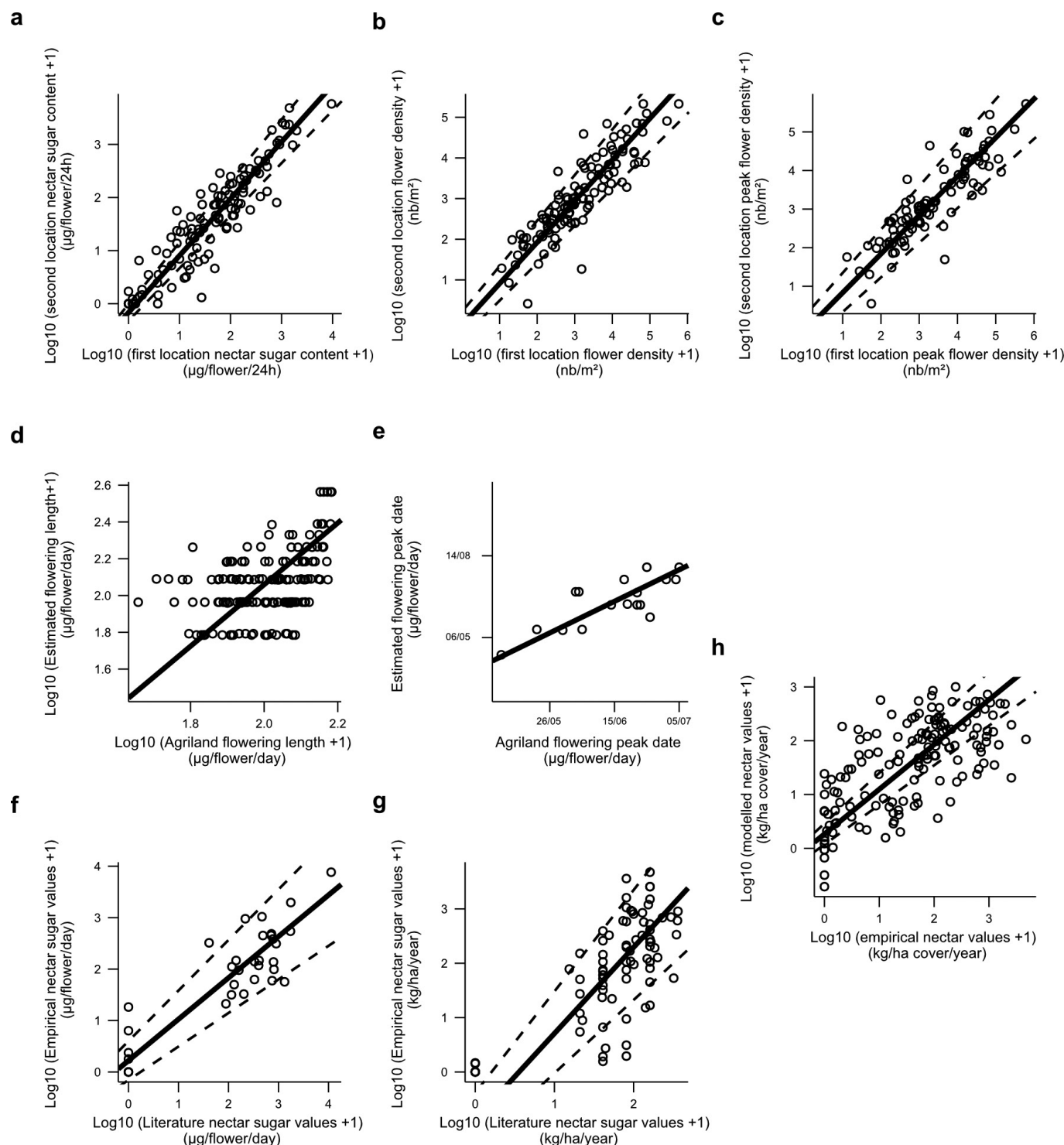
Baude et al.  
What? Changes in nectar provision at the national scale  
Where? Great Britain  
Database? Plant species nectar productivity combined to national vegetation and land covers surveys



1930 1950 1970 1990 2010

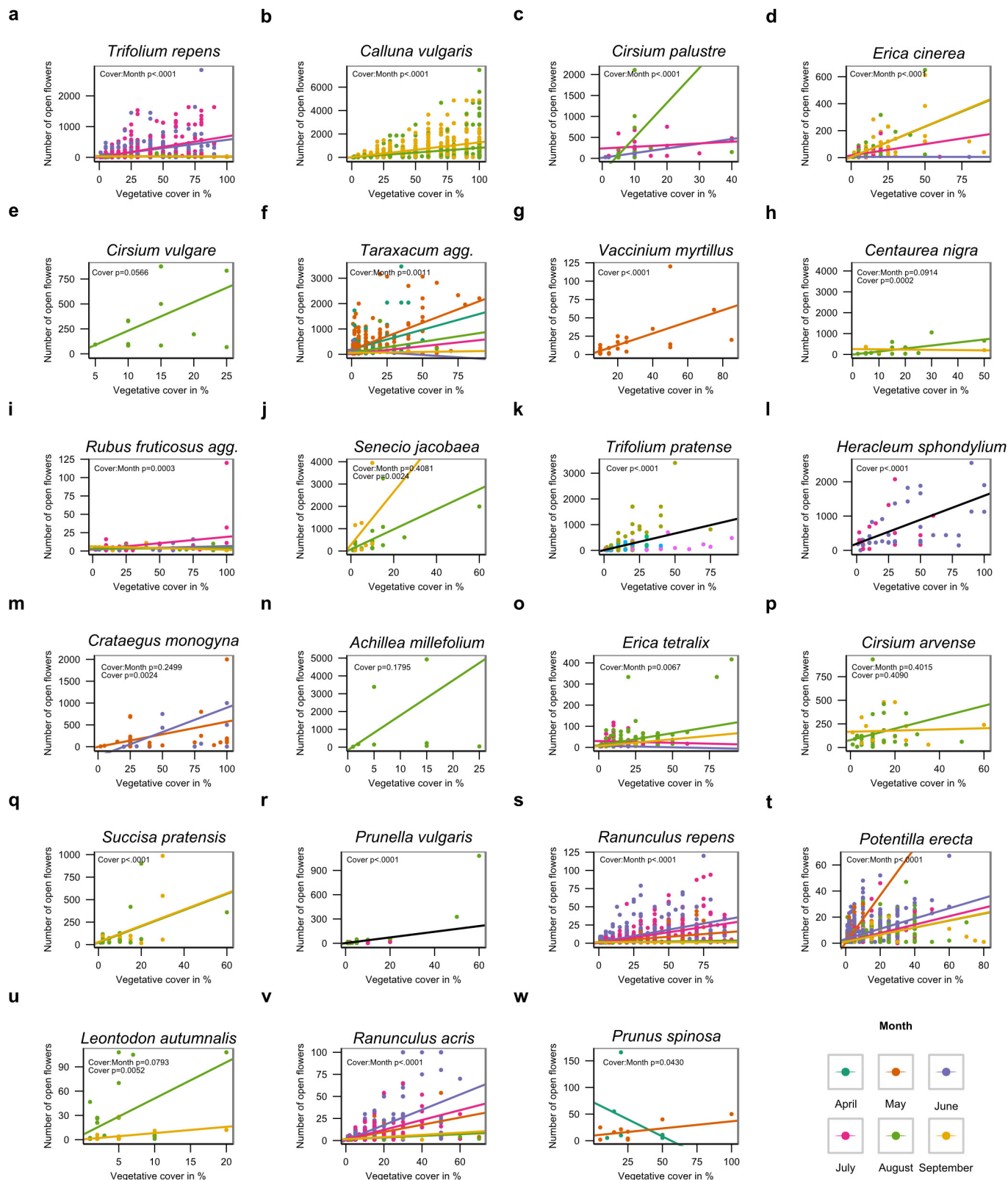
**Extended Data Figure 5 | Historical timeline in changes in nectar resources and flower-visiting insects in Great Britain.** Historical periods with the greatest negative changes in nectar resources (in percentage change in England and Wales (E&W) and Great Britain (GB)) and flower-visiting insects (in percentage change or in species extinction per decade (sp./dec.)) are indicated in red, those with intermediate changes

are in orange, and those with the lowest (or even reversing) changes are in green. Main historical trends from this study (Baude *et al.*) are presented in regard to those described in Carvalho *et al.*<sup>10</sup> and Ollerton *et al.*<sup>9</sup> studies. The white chevron indicates a provisional extinction rate that needs to be confirmed on a 20 year period of time (see Supplementary Information from Ollerton *et al.*<sup>9</sup>).



**Extended Data Figure 6 | Validity of the data sets.** **a**, Major axis linear regression of  $\log_{10}(x+1)$  nectar values per flower obtained in the second location against those obtained in the first one. **b**, Major axis linear regression of  $\log_{10}(x+1)$  flower density values obtained in the second location against those obtained in the first one. **c**, Major axis linear regression of  $\log_{10}(x+1)$  peak flower density values obtained in the second location against those obtained in the first one. **d**, Standardized major axis regression of the  $\log_{10}(x+1)$  length of the flowering period used for analyses with those derived from IPI AgriLand floral transects (unpublished data). **e**, Standardized major axis regression of peak date of flowering season used for analyses with those derived from IPI AgriLand floral transects (unpublished data). **f**, Major axis linear regression

performed on the  $\log_{10}(x+1)$  empirical (empirical data set) and published nectar values (literature data set from Raine and Chittka<sup>40</sup>) at the flower scale. **g**, Standardized major axis linear regression performed on the  $\log_{10}(x+1)$  empirical (empirical data set) and published nectar values (literature data set, see Supplementary Table 13 for references) at the vegetative scale. **h**, Standardized major axis linear regression performed on the  $\log_{10}(x+1)$  empirical and modelled nectar values generated by a leave-one-out approach. Estimated values and 95% confidence intervals (solid and dashed lines, respectively) of all equations are derived from (standardized) major axis regression (ma and sma function from 'smatr' package in R<sup>36</sup>; see Supplementary Information for details).



**Extended Data Figure 7 | Flower number and vegetative cover relationships.** Linear regressions between the number of open flowers counted in a quadrat of  $0.5 \text{ m}^2$  according to the vegetative cover of the focus species in the quadrat (in %). Data are extracted from IPI AgriLand floral transects survey in 2012 (unpublished data) for 23 out of the 35 main nectar contributing species (panels a–w). The number of flowers was analysed according to the vegetative cover ('Cover'), the month of the survey ('Month') and the interaction between these two terms

('Cover:Month') using negative binomial generalized linear models (see Supplementary Information for details). Coloured lines represent the linear regression between flower abundance and vegetative cover for each month of the survey. Black lines represent the overall linear regression between flower abundance and vegetative cover when the 'Month' covariate cannot be included in the model. Line equations were derived from statistical intercept and slope estimates.



**Extended Data Table 1 | Type III ANOVA (F-tests) results for annual nectar productivity, species nectar diversity and functional nectar diversity****a**

Response variable	Effect	df	F value	P-value
Nectar productivity	Habitat	10	69.643	<.0001
Species nectar diversity	Habitat	10	19.923	<.0001
Functional nectar diversity	Habitat	10	24.150	<.0001

**b**

Response variable	Effect	df	F value	P-value
Nectar productivity	Habitat	10	75.081	<.0001
	Location	1	0.560	0.455
	Habitat:Location	10	63.519	<.0001
Species nectar diversity	Habitat	10	22.061	<.0001
	Location	1	0.147	0.701
	Habitat:Location	10	10.396	<.0001
Functional nectar diversity	Habitat	10	23.677	<.0001
	Location	1	2.158	0.142
	Habitat:Location	10	15.810	<.0001

**c**

Response variable	Effect	df	F value	P-value
Nectar productivity	Habitat	10	26.860	<.0001
	Year	3	1.473	0.220
	Habitat:Year	30	1.793	0.005
Species nectar diversity	Habitat	10	5.137	<.0001
	Year	3	2.600	0.050
	Habitat:Year	30	2.523	<.0001
Functional nectar diversity	Habitat	10	3.517	0.0001
	Year	3	1.987	0.114
	Habitat:Year	30	1.725	0.009

**a**, 2007 values according to habitat. The linear mixed effect models were performed on data from 2,576 nonlinear plots surveyed in 2007. **b**, 2007 values according to habitat and location. The linear mixed effect models were performed on data from 4,527 plots (2,576 nonlinear plots and 1,951 linear plots) surveyed in 2007. **c**, 1978–2007 values according to habitat and year. The linear mixed effect models were performed on data from 768 shared plots surveyed in 1978, 1990, 1998 and 2007. The annual nectar productivity was systematically  $\log_{10}(x+1)$  transformed. See Supplementary Table 1 and Supplementary Table 3 for sample sizes.

Extended Data Table 2 | Flower morphology and flower-visiting insects of the four main nectar-providing species

	Depth of nectar tube		Width of nectar tube		Number of visiting insect species					Frequent visiting insect species	Number of sources
	mean	sem	mean	sem	All	Diptera	Hymeno	Lepido	Coleo		
<i>Trifolium repens</i>	4.84	0.19	1.36	0.04	54	22	16 (13 species of <i>Bombus</i> )	8	8	<i>Bombus pascuorum</i> , <i>Bombus lucorum/terrestris</i> , <i>Bombus lapidarius</i>	21
<i>Calluna vulgaris</i>	2.23	0.10	1.93	0.07	139	96	29 (9 species of <i>Bombus</i> )	13	1	<i>Bombus lucorum/terrestris</i> , <i>Bombus pascuorum</i> , <i>Apis mellifera</i> , <i>Bombus jonellus</i>	9
<i>Cirsium palustre</i>	3.63	0.07	1.42	0.07	12	5	7 (6 species of <i>Bombus</i> )	0	0	<i>Bombus pascuorum</i> , <i>Bombus lucorum/terrestris</i> , <i>Bombus pratorum</i>	6
<i>Erica cinerea</i>	5.81	0.11	1.67	0.06	49	19	27 (10 species of <i>Bombus</i> )	2	1	<i>Bombus jonellus</i> , <i>Bombus lucorum/terrestris</i> , <i>Bombus pascorum</i>	6

Flower morphology parameters (mean and standard error for depth and width of flower tubes) were measured on 20–40 flowers per species in the field. Flower-visiting insects were listed from plant–insect visiting networks from J.M. (published and unpublished data to which recorded interactions from a review of literature have been added (see Supplementary Table 12 for reference list)).

**Extended Data Table 3 | Agri-environment schemes and linear features: nectar productivity and provision in England in 2007****a**

Option	Option code	Mean nectar productivity (kg of sugars/ha/year)	England land cover (000s ha)	England nectar provision (000 000s kg of sugars/year)
Wild bird seed mixture	EF2/HF2	56.00	2.97	0.17
Enhanced grass buffer strip	HE10	166.80	0.62	0.10
Nectar flower mixture	HF4/HF4	244.00	1.61	0.39
Haymaking supplement	HK18	18.60	1.12	0.02
Species-rich semi-natural grassland	HK6/7/8	31.90	2.77	0.09

**b**

Linear features	Linear code	England mean nectar productivity (kg of sugars/ha/year)	England land cover (000s ha)	England nectar provision (000 000s kg of sugars/year)
Hedgerows	H	341.59	40.20	13.73
Watersides	S	60.97	/	/
Road verges	R	60.63	/	/

**a.** Mean nectar productivity values of agri-environment schemes were estimated from our nectar productivity database combined with flower counts in these options. Areas of options providing floral resources in the English agri-environment scheme 'Environmental Stewardship' were extracted for 2007 from data supplied by Natural England<sup>38,39</sup>. **b.** Mean nectar productivity values of linear features correspond to back-transformed ( $10^x - 1$ ) estimates of the linear mixed model fitted on  $\log_{10}(x + 1)$  nectar productivity of all Countryside Survey linear plots surveyed in England in 2007. National areas of hedgerows were estimated from the length given in Countryside Survey 2007 for England<sup>35</sup> and assuming a 1 m width.

# Xenacoelomorpha is the sister group to Nephrozoa

Johanna Taylor Cannon<sup>1</sup>, Bruno Cossermelli Vellutini<sup>2</sup>, Julian Smith III<sup>3</sup>, Fredrik Ronquist<sup>1</sup>, Ulf Jondelius<sup>1</sup> & Andreas Hejnol<sup>2</sup>

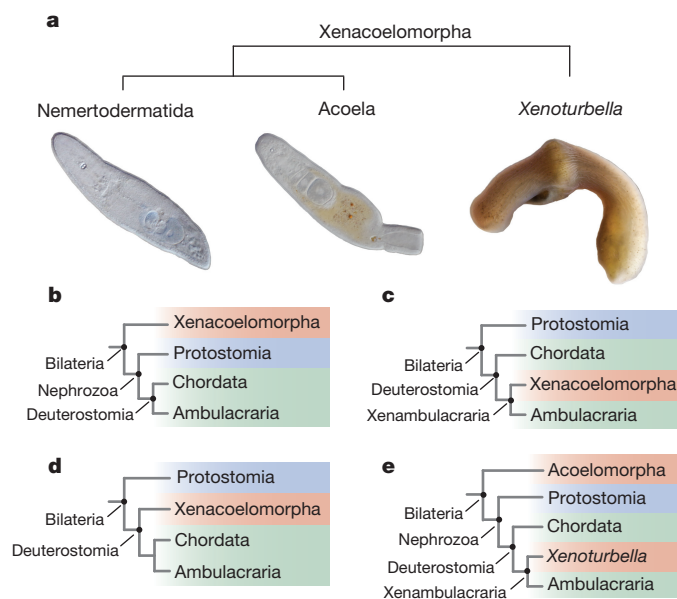
The position of Xenacoelomorpha in the tree of life remains a major unresolved question in the study of deep animal relationships<sup>1</sup>. Xenacoelomorpha, comprising Acoela, Nemertodermatida, and *Xenoturbella*, are bilaterally symmetrical marine worms that lack several features common to most other bilaterians, for example an anus, nephridia, and a circulatory system. Two conflicting hypotheses are under debate: Xenacoelomorpha is the sister group to all remaining Bilateria (= Nephrozoa, namely protostomes and deuterostomes)<sup>2,3</sup> or is a clade inside Deuterostomia<sup>4</sup>. Thus, determining the phylogenetic position of this clade is pivotal for understanding the early evolution of bilaterian features, or as a case of drastic secondary loss of complexity. Here we show robust phylogenomic support for Xenacoelomorpha as the sister taxon of Nephrozoa. Our phylogenetic analyses, based on 11 novel xenacoelomorph transcriptomes and using different models of evolution under maximum likelihood and Bayesian inference analyses, strongly corroborate this result. Rigorous testing of 25 experimental data sets designed to exclude data partitions and taxa potentially prone to reconstruction biases indicates that long-branch attraction, saturation, and missing data do not influence these results. The sister group relationship between Nephrozoa and Xenacoelomorpha supported by our phylogenomic analyses implies that the last common ancestor of bilaterians was probably a benthic, ciliated acoelomate worm with a single opening into an epithelial gut, and that excretory organs, coelomic cavities, and nerve cords evolved after xenacoelomorphs separated from the stem lineage of Nephrozoa.

Acoela have an essential role in hypotheses of bilaterian body plan evolution<sup>5</sup>. Acoels have been compared to cnidarian planula larvae because they possess characters such as a blind gut, a net-like nervous system, and they lack nephridia. However, they also share apomorphies with Bilateria such as bilateral symmetry and a mesodermal germ layer that gives rise to circular and longitudinal muscles. Classic systematics placed acoels in Platyhelminthes<sup>6</sup>, or as a separate early bilaterian lineage<sup>7,8</sup>. When nucleotide sequence data became available, Acoela were placed as the sister group of Nephrozoa<sup>9</sup>. Nemertodermatida were originally classified within Acoela, but were soon recognized as a separate clade on morphological grounds<sup>10</sup>. Subsequently, nucleotide sequence data fuelled a debate on whether nemertodermatids and acoels form a monophyletic group, the Acoelomorpha, or if nemertodermatids and acoels are independent early bilaterian lineages as suggested by several studies, for example refs 11 and 12. The enigmatic *Xenoturbella* was first placed together with Acoela and Nemertodermatida<sup>13,14</sup>, then an ultrastructural appraisal supported its position as sister group of all other bilaterians<sup>15</sup>. The first molecular study suggested *Xenoturbella* to be closely related to molluscs<sup>16</sup>, whereas other analyses proposed a deuterostome affiliation<sup>17,18</sup>. Recent analyses of molecular data reunited *Xenoturbella* with acoels and nemertodermatids<sup>2–4</sup> to form a clade called Xenacoelomorpha (Fig. 1a).

Current conflicting hypotheses suggest that Xenacoelomorpha are the sister group of Deuterostomia<sup>4</sup>, are nested within Deuterostomia<sup>4</sup>, are the sister group of Nephrozoa<sup>2,3</sup>, or are polyphyletic, with *Xenoturbella* included within Deuterostomia and the Acoelomorpha

as sister taxon to remaining Bilateria<sup>19</sup> (Fig. 1b–e). The deuterostome affiliation derives support from three lines of evidence<sup>4</sup>: an analysis of mitochondrial gene sequences, microRNA complements, and a phylogenomic data set. Analyses of mitochondrial genes recovered *Xenoturbella* within deuterostomes<sup>18</sup>. However, limited mitochondrial data (typically ~16 kilobase total nucleotides, 13 protein-coding genes) are less efficient in recovering higher-level animal relationships than phylogenomic approaches, especially in long-branching taxa<sup>1</sup>. The one complete and few partial mitochondrial genomes for acoelomorphs are highly divergent in terms of both gene order and nucleotide sequence<sup>19,20</sup>. Analyses of new complete mitochondrial genomes of *Xenoturbella* spp. do not support any phylogenetic hypothesis for this taxon<sup>21</sup>. Ref. 4 proposes that microRNA data support Xenacoelomorpha within the deuterostomes; however, microRNA distribution is better explained by a sister relationship between Xenacoelomorpha and Nephrozoa both under parsimony<sup>4,22</sup> and under Bayesian inference<sup>22</sup>.

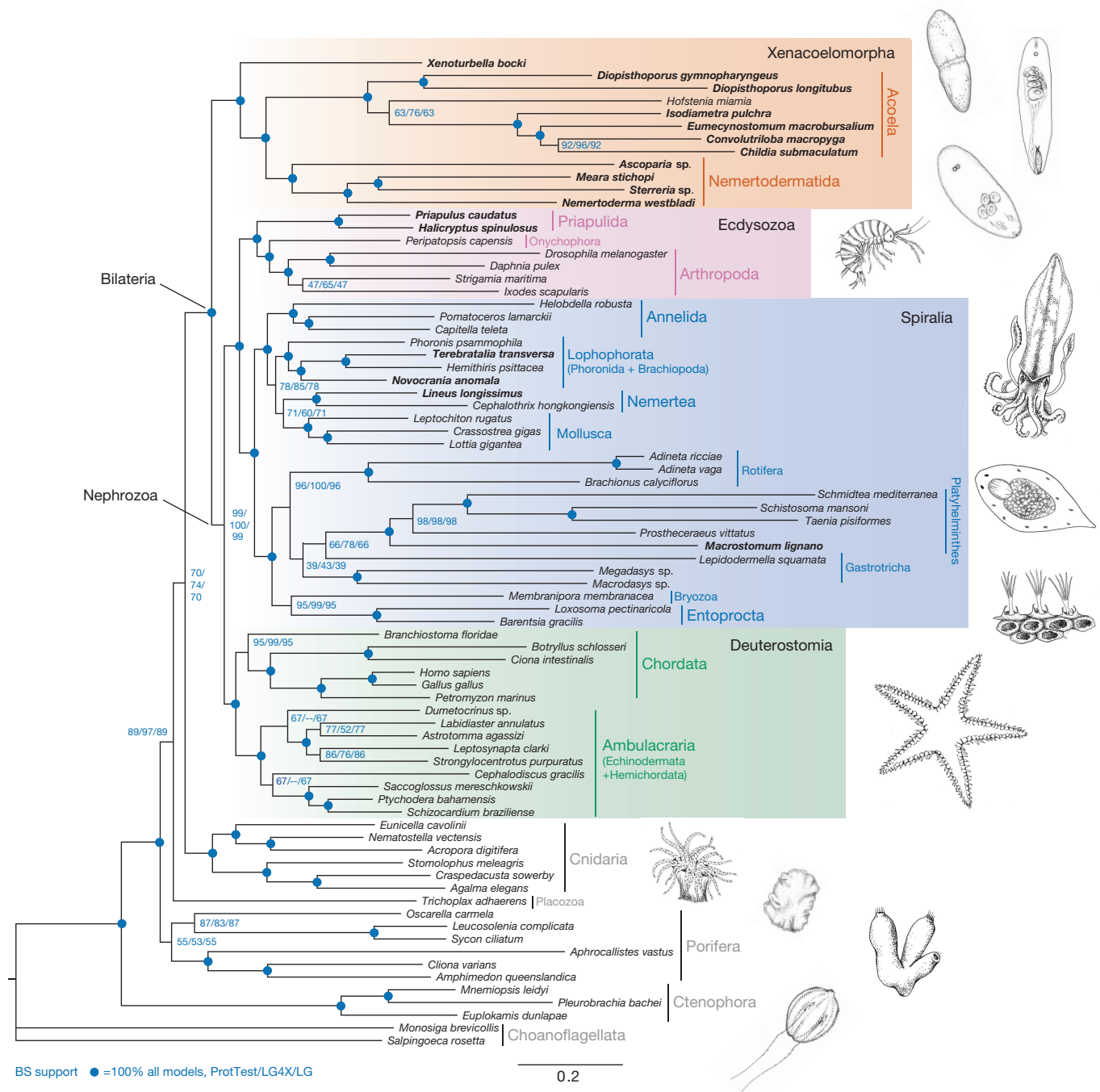
Phylogenomic analyses recovering xenacoelomorph taxa within Deuterostomia show branching patterns that differ significantly



**Figure 1 | Phylogenetic hypotheses concerning Xenacoelomorpha from previous molecular studies.** **a**, Relationships among Xenacoelomorpha. *Xenoturbella* is sister to Acoelomorpha (Acoela + Nemertodermatida). Illustrated species from left to right: *Flagellophora apelti*, *Diopisthoporus psammophilus*, *X. bocki*. **b**, Xenacoelomorpha is sister taxon to Nephrozoa (phylogenomic analyses<sup>2,3</sup>). **c**, Xenacoelomorpha is sister taxon to Ambulacraria within deuterostomes (phylogenomic analyses<sup>4</sup>). **d**, Xenacoelomorpha is sister taxon to Ambulacraria + Chordata (mitochondrial protein analyses<sup>4,19</sup>). **e**, *Xenoturbella* is within Deuterostomia, while Acoelomorpha form two separate clades outside Nephrozoa (molecular systematic analyses<sup>11</sup>), or its sister group (some mitochondrial protein analyses<sup>19</sup>). Colours in **b–e** indicate Xenacoelomorpha (red), Protostomia (blue), Deuterostomia (green).

<sup>1</sup>Naturhistoriska Riksmuseet, PO Box 50007, SE-104 05 Stockholm, Sweden. <sup>2</sup>Sars International Centre for Marine Molecular Biology, University of Bergen, Thormøhlensgate 55, 5008 Bergen, Norway. <sup>3</sup>Department of Biology, Winthrop University, 701 Oakland Avenue, Rock Hill, South Carolina 29733, USA.





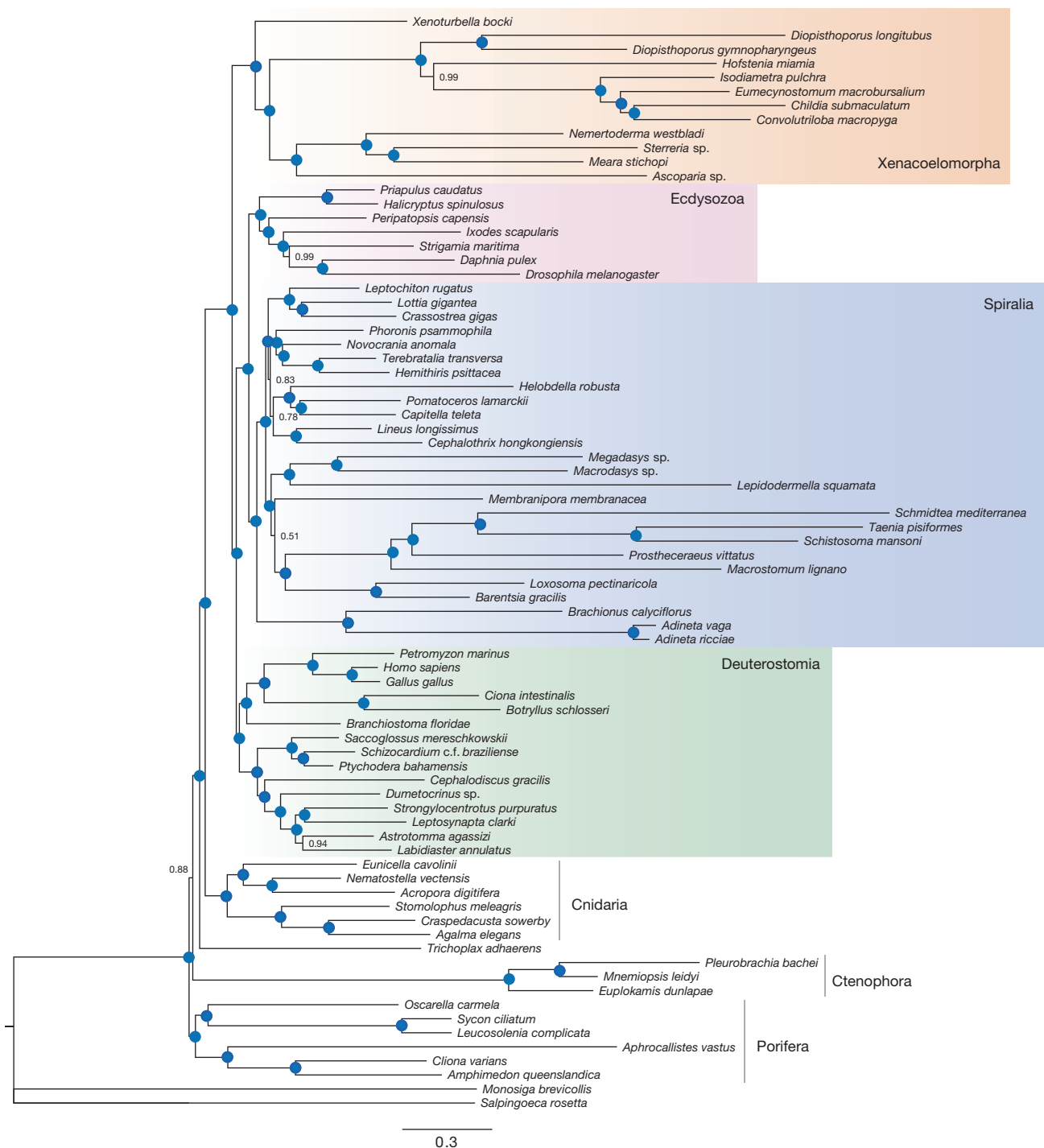
**Figure 2 | Maximum likelihood topology of metazoan relationships inferred from 212 genes.** Maximum likelihood tree is shown as inferred using the best-fitting amino-acid substitution model for each gene. Bootstrap support values from analyses inferred under alternative models of amino-acid substitution are indicated at the nodes (best-fitting

between alternative models of evolution<sup>4</sup>. Conflicting results in studies that used the same expressed sequence tag data for xenacoelomorphs<sup>2,4</sup> suggest some degree of model misspecification, missing data generating positively misleading signal, or long-branch attraction (LBA) in either or both of these studies. Testing of hypotheses under alternative models of evolution, data set partitioning, and taxon selection schemes can identify possible weaknesses of a data set. Here, we use this approach to test the phylogenetic position of Acoela, Nemertodermatida, and *Xenoturbella*.

Novel Illumina RNaseq data were collected for six acoel species, four nemertodermatids, *Xenoturbella bocki*, and six additional

model for each orthologous group selected by ProtTest/LG4X across all partitions/LG + I +  $\Gamma$  across all partitions, 100 bootstrap replicates). Filled blue circles represent 100% bootstrap support under all models of evolution. Species indicated in bold are new transcriptomes published with this study.

diverse metazoans (Supplementary Table 1). Acoel and nemertodermatid species were selected to broadly represent the diversity of these two clades, including two representatives of the earliest-branching clade of Acoela, Diopisthoporidae<sup>23</sup>. With the exception of *Hofstenia miamia* in ref. 3, previous phylogenomic analyses of acoels have included only representatives of Convolutidae and Isodiametridae, which possess several highly derived morphological characters. Our data sets include 76 diverse metazoan taxa and 2 choanoflagellate outgroups (Supplementary Table 1). Our primary data set consists of 212 orthologous groups, 44,896 amino-acid positions, and 31% missing data (Extended Data Table 1).



**Figure 3 | Bayesian inference topology of metazoan relationships inferred from 212 genes under the CAT + GTR +  $\Gamma$  model.** Filled blue circles indicate posterior probabilities of 1.0. Shown is the majority rule consensus tree of two independent chains of > 17,000 cycles each and burn-in of 5,000 cycles. Convergence of the two chains was indicated by

a 'maxdiff' value of 0.25. Position of Xenacoelomorpha was unchanged in two additional independent chains, which did not converge with the chains shown above owing to alternative positions of *Trichoplax adhaerens* and *Membranipora membranacea*.

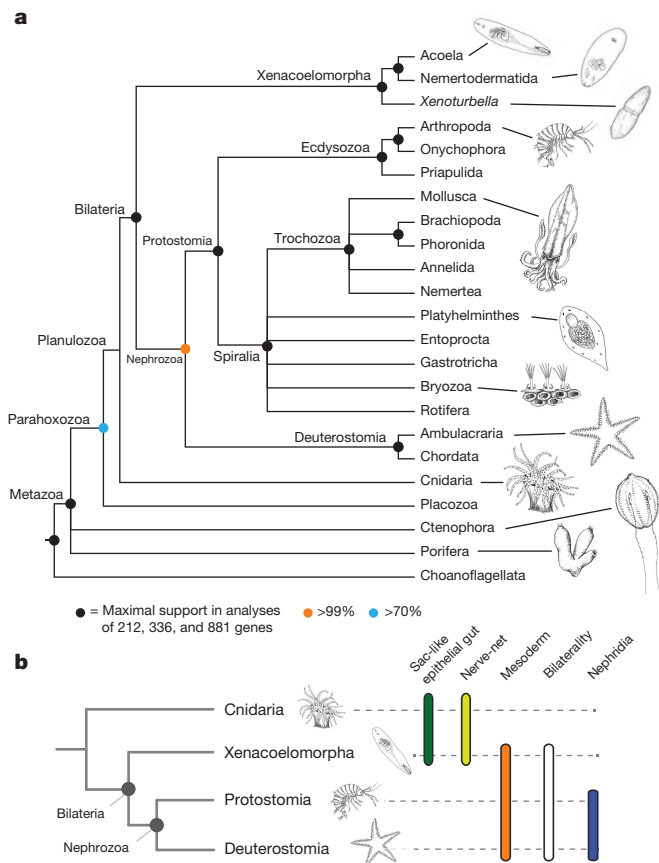
Sequences were taken entirely from Illumina transcriptomes or predicted transcripts from genomic data. Gene occupancy per taxon ranged from 100% for *Homo sapiens* and *Drosophila melanogaster* to 8% for the nemertodermatid *Sterreria* sp., with median per-taxon gene occupancy of 90% and an average of 80% (Supplementary Table 2). Notably, gene coverage for key taxa is enhanced over previous phylogenomic analyses: *X. bocki*, six acoels, and two nemertodermatids have > 90% gene occupancy in our 212 orthologous group data set, whereas the best represented acoelomorph terminal in ref. 4 had an occupancy of 63%.

Maximum likelihood analyses were conducted under the best-fitting model for each individual gene partition, or the LG model, or the LG4X model<sup>24</sup> over each independent partition. The LG4X model is composed of four substitution matrixes designed to improve modelling of site heterogeneity<sup>24</sup>. Bayesian analyses were conducted with the site-heterogeneous CAT + GTR +  $\Gamma$  model and GTR +  $\Gamma$ . To further validate the robustness of our results to variations in substitution model specification, we performed Bayesian inference analyses under an independent substitution model using a back-translated nucleotide data set derived from our amino-acid alignment. To test whether any

particular taxon was biasing our analyses owing to artefacts such as LBA, we conducted a series of taxon-pruning experiments. Additional data sets were analysed that minimized missing data, excluded taxa and individual genes identified to be potentially more subject to LBA artefacts, and genes or positions that were more saturated. Using our standard pipeline, for the best-sampled 56 taxa, we also generated a data set with 336 orthologous groups, 81,451 amino acids, and 11% missing data. Lastly, using an independent pipeline for orthologous gene selection, we generated a set of 881 orthologous groups. This larger data set contained 77 operational taxonomic units, 337,954 amino-acid positions, and 63% matrix occupancy. In all, we generated 25 unique data matrixes to address the robustness of phylogenetic signal and sensitivity of our results to parameter changes (Extended Data Table 1).

Our analyses consistently supported monophyletic Xenacoelomorpha as sister group of Nephrozoa (Figs 2–4, Extended Data Figs 1–4 and Extended Data Table 1). Within Xenacoelomorpha, *Xenoturbella* is the sister taxon of Acoela + Nemertodermatida. Maximum likelihood analyses under all models (Fig. 2), Bayesian analyses under the site-heterogeneous CAT + GTR +  $\Gamma$  model (Fig. 3), as well as analyses of back-translated nucleotides (Extended Data Fig. 5) all recover this topology. We found no evidence of LBA influencing the position of Xenacoelomorpha or any other group in the tree. Differing outgroup schemes do not affect the position of Xenacoelomorpha (Supplementary Figs 4–9); neither does exclusion of taxa or genes more subject to LBA (Supplementary Figs 14–17). Monophyletic Deuterostomia (excluding *Xenoturbella*), Ecdysozoa, and Spiralia are robustly recovered, with Ctenophora as the earliest branching metazoan in all maximum likelihood analyses, while Porifera holds this position in Bayesian analyses under the CAT + GTR +  $\Gamma$  model (Fig. 3). Taxon-exclusion analyses, where Acoelomorpha alone (Supplementary Fig. 1) or *Xenoturbella* alone (Extended Data Fig. 3) were included, recovered these taxa as the first branch of Bilateria. Approximately unbiased tests strongly reject the alternative hypothesis constraining Xenacoelomorpha within Deuterostomia. Leaf stability indices for all taxa in the primary 212 orthologous group analysis were > 97% (Supplementary Table 2), suggesting that improved matrix and taxon coverage in our analyses had a positive effect on overall taxon stability compared with ref. 25, where both included acoels had leaf stability indices of 78%. In our own calculations of leaf stability index from the data set of ref. 4, the six representative xenacoelomorph species have the six lowest leaf stabilities of all included taxa, ranging from 88% to 79% (Supplementary Table 3).

To assess gene conflict, we conducted decomposition analyses using ASTRAL<sup>26</sup>, which calculates the species tree that agrees with the largest number of quartets derived from each gene tree and their respective bootstrap replicates (Extended Data Fig. 4). This analysis finds strong support for the position of Xenacoelomorpha (bootstrap 99%). Refs 27 and 28 pointed to issues with incongruence in phylogenomic analyses of ribosomal protein genes versus other protein-coding genes. Notably, in our 212 orthologous group set, only five ribosomal protein genes were retained after screening for paralogous groups. To investigate if this gene class may have biased previous results, we generated an additional data matrix composed of 52 ribosomal protein genes that passed through our other filters for gene length and taxon presence. In maximum likelihood analyses of this data set, Xenacoelomorpha, Acoelomorpha, Nemertodermatida, Deuterostomia and Spiralia are all non-monophyletic (Supplementary Fig. 21). Ribosomal protein genes are heavily represented in the xenacoelomorph data in previous studies, comprising > 50% of the gene occupancy in most cases. Gene partition information was not made available for the study proposing a deuterostome position for Xenacoelomorpha<sup>4</sup>, so re-analysis of the data without ribosomal protein genes was not possible. We suggest that insufficient data for key taxa and a reliance on ribosomal protein genes were biasing the results, causing Xenacoelomorpha to group within Deuterostomia.



**Figure 4 | Summary of metazoan relationships as inferred in this study.** **a**, Summary of phylogenomic results based on analyses of 212, 336, and 881 genes. Xenacoelomorpha is a monophyletic clade sister to Nephrozoa with > 99% support in all analyses. **b**, Interrelationships among four major animal clades, Cnidaria, Xenacoelomorpha, Protostomia, and Deuterostomia, with selected morphological characters mapped onto the tree as ancestral states for each of the four clades.

Within Xenacoelomorpha, morphological complexity differs among the three groups, as should be expected in a clade of the same age as Nephrozoa. The simplest organization is evident in *Xenoturbella*, with a sac-like epithelial gut opening to a simple mouth, a basiepidermal nervous system, and no gonopores or secondary reproductive organs<sup>13</sup>. Nemertodermatida also have an epithelial gut, but the mouth appears to be a transient structure<sup>10</sup>. Furthermore, the position and anatomy of the nervous system and the male copulatory organ are variable. The more than 400 nominal species of Acoela (compared with 18 nemertodermatids and 5 *Xenoturbella* species) exhibit considerable morphological variation: acoels have no intestinal lumen although a mouth opening and sometimes a pharynx is present<sup>23</sup>. The nervous system is highly variable, there are one or two gonopores, and often accessory reproductive organs<sup>23</sup>. The morphological evolution that occurred within Xenacoelomorpha provides an interesting parallel case to Nephrozoa.

The sister group relationship between Xenacoelomorpha and Nephrozoa allows us to infer the order in which bilaterian features were evolved<sup>12,29</sup>. The bilaterian ancestor was probably a soft-bodied, small ciliated benthic worm<sup>5,23,29,30</sup>. Mesoderm and body axis were established before the split between Xenacoelomorpha and Nephrozoa, whereas nephridia evolved in the stem lineage of nephrozoans (Fig. 4). Centralization of the nervous system appears to have evolved in parallel in the Xenacoelomorpha and Nephrozoa. Further investigations of the genomic architecture and biology of xenacoelomorphs will provide insights into molecular, developmental, and cellular building blocks used for evolving complex animal body plans and organ systems.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 19 September; accepted 7 December 2015.**

1. Dunn, C. W., Giribet, G., Edgecombe, G. D. & Hejnol, A. Animal phylogeny and its evolutionary implications. *Annu. Rev. Ecol. Syst.* **45**, 371–395 (2014).
2. Hejnol, A. *et al.* Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc. R. Soc. B* **276**, 4261–4270 (2009).
3. Srivastava, M., Mazza-Curli, K. L., van Wolfswinkel, J. C. & Reddien, P. W. Whole-body acoel regeneration is controlled by Wnt and Bmp-Admp signaling. *Curr. Biol.* **24**, 1107–1113 (2014).
4. Philippe, H. *et al.* Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. *Nature* **470**, 255–258 (2011).
5. Nielsen, C. *Animal Evolution: Interrelationships of the Living Phyla* (Oxford Univ. Press, 2012).
6. Ehlers, U. *Das phylogenetische System der Plathelminthes* (G. Fischer, 1985).
7. Smith, J. P. S., III, Tyler, S. & Rieger, R. M. Is the Turbellaria polyphyletic? *Hydrobiologia* **132**, 13–21 (1986).
8. Haszprunar, G. Plathelminthes and Plathelminthomorpha — paraphyletic taxa. *J. Zool. Syst. Evol. Res.* **34**, 41–48 (1996).
9. Ruiz-Trillo, I., Riutort, M., Littlewood, D. T. J., Herniou, E. A. & Baguña, J. Acoel flatworms: earliest extant bilaterian metazoans, not members of Platyhelminthes. *Science* **283**, 1919–1923 (1999).
10. Steinböck, O. Ergebnisse einer von E. Reisinger & O. Steinböck mit Hilfe des Rask-Örsted fonds durchgeführten Reise in Grönland 1926. 2. *Nemertoderma bathycola* nov. gen. nov. spec., eine eigenartige Turbellarie aus der Tiefe der Diskobay: nebst einem Beitrag zur Kenntnis des Nemertinenepithels. *Vidensk. Medd. Dan. Naturhist. Foren.* **90**, 47–84 (1930).
11. Paps, J., Baguña, J. & Riutort, M. Bilaterian phylogeny: a broad sampling of 13 nuclear genes provides a new Lophotrochozoa phylogeny and supports a paraphyletic basal acoelomorpha. *Mol. Biol. Evol.* **26**, 2397–2406 (2009).
12. Jondelius, U., Ruiz-Trillo, I., Baguña, J. & Riutort, M. The Nemertodermatida are basal bilaterians and not members of the Platyhelminthes. *Zool. Scr.* **31**, 201–215 (2002).
13. Westblad, E. *Xenoturbella bocki* n.g. n.sp. a peculiar, primitive turbellarian type. *Ark. Zool.* **1**, 3–29 (1949).
14. Franzén, Å. & Afzelius, B. A. The ciliated epidermis of *Xenoturbella bocki* (Platyhelminthes, Xenoturbellida) with some phylogenetic considerations. *Zool. Scr.* **16**, 9–17 (1987).
15. Ehlers, U. & Sopott-Ehlers, B. Ultrastructure of the subepidermal musculature of *Xenoturbella bocki*, the adelphotaxon of the Bilateria. *Zoomorphology* **117**, 71–79 (1997).
16. Norén, M. & Jondelius, U. *Xenoturbella*'s molluscan relatives... *Nature* **390**, 31–32 (1997).
17. Bourlat, S. J. *et al.* Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85–88 (2006).
18. Bourlat, S. J., Rota-Stabelli, O., Lanfear, R. & Telford, M. J. The mitochondrial genome structure of *Xenoturbella bocki* (phylum Xenoturbellida) is ancestral within the deuterostomes. *BMC Evol. Biol.* **9**, 107 (2009).
19. Mwinzi, A. *et al.* The phylogenetic position of Acoela as revealed by the complete mitochondrial genome of *Symsagittifera roscoffensis*. *BMC Evol. Biol.* **10**, 309 (2010).
20. Ruiz-Trillo, I., Riutort, M., Fourcade, H. M., Baguña, J. & Boore, J. L. Mitochondrial genome data support the basal position of Acoelomorpha and the polyphyly of the Platyhelminthes. *Mol. Phylogenet. Evol.* **33**, 321–332 (2004).
21. Rouse, G., Wilson, N. G., Carvajal, J. I. & Vrijenhoek, R. C. New deep-sea species of *Xenoturbella* and the position of Xenacoelomorpha. *Nature* <http://dx.doi.org/10.1038/nature16545> (this issue).
22. Thomson, R. C., Plachetzki, D. C., Mahler, D. L. & Moore, B. R. A critical appraisal of the use of microRNA data in phylogenetics. *Proc. Natl Acad. Sci. USA* **111**, E3659–E3668 (2014).
23. Jondelius, U., Wallberg, A., Hooge, M. & Raikova, O. I. How the worm got its pharynx: phylogeny, classification and Bayesian assessment of character evolution in Acoela. *Syst. Biol.* **60**, 845–871 (2011).
24. Le, S. Q., Dang, C. C. & Gascuel, O. Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Mol. Biol. Evol.* **29**, 2921–2936 (2012).
25. Dunn, C. W. *et al.* Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–749 (2008).
26. Mirarab, S. *et al.* ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–i548 (2014).
27. Bleidorn, C. *et al.* On the phylogenetic position of Myzostomida: can 77 genes get it wrong? *BMC Evol. Biol.* **9**, 150 (2009).
28. Whelan, N. V., Kocot, K. M., Moroz, L. L. & Halanych, K. M. Error, signal, and the placement of Ctenophora sister to all other animals. *Proc. Natl Acad. Sci. USA* **112**, 5773–5778 (2015).
29. Hejnol, A. & Martindale, M. Q. Acoel development supports a simple planula-like urbilaterian. *Phil. Trans. R. Soc. B* **363**, 1493–1501 (2008).
30. Laumer, C. E. *et al.* Spiralian phylogeny informs the evolution of microscopic lineages. *Curr. Biol.* **25**, 2000–2006 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The Swedish Research Council provided funding for U.J. and J.T.C. (grant 2012-3913) and F.R. (grant 2014-5901). A.H. received support from the Sars Core budget and Marie Curie Innovative Training Networks 'NEPTUNE' (FP7-PEOPLE-2012-ITN 317172) and FP7-PEOPLE-2009-RG 256450. We thank N. Lartillot and K. Kocot for discussions. Hejnol laboratory members K. Pang and A. Børve assisted with RNA extraction; A. Boddington, J. Bengtson and A. Elde assisted with culture for *Isodiametra pulchra* and *Convolutriloba macropyga*. Thanks to W. Sterrer for collection of *Sterreria* sp. and *Ascoparia* sp., and to R. Janssen for finding *X. bocki*. The Sven Lovén Centre of Marine Sciences Kristineberg, University of Gothenburg, and the Interuniversity Institute of Marine Sciences in Eilat provided logistical support for field collection. S. Baldauf assisted with laboratory space and resources for complementary DNA synthesis. We thank K. Larsson for the original illustrations. Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC). Transcriptome assembly, data set construction, RAXML and PhyloBayes analyses were performed using resources provided through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project b2013077, and MrBayes analyses were run under project snic2014-1-323.

**Author Contributions** J.T.C., U.J., B.C.V., and A.H. conceived and designed the study. U.J. and A.H. collected several specimens and J.S. III collected *Diopisthoporus gymnopharyngeus* specimens. J.T.C. and B.C.V. performed molecular work and RNA sequencing assembly. J.T.C. assembled the datasets and performed phylogenetic analyses. F.R. conducted Bayesian phylogenetic analyses using MrBayes. All authors contributed to writing the manuscript.

**Author Information** Sequence data have been deposited in the NCBI Sequence Read Archive under BioProject PRJNA295688. Data matrices and trees from this study are available from the Dryad Digital Repository (<http://datadryad.org>) under DOI 10.5061/dryad.493b7. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.T.C. (joie.cannon@gmail.com) or A.H. (andreas.hejnol@uib.no).



## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Molecular methods and sequencing.** We generated novel RNA-seq data from six acoels, four nemertodermatids, *X. bocki*, and six additional diverse metazoans (Supplementary Table 1). Total RNA was extracted from fresh or RNAlater (Ambion) preserved specimens using TRI Reagent Solution (Ambion) or the RNeasy Micro Kit (Qiagen), prepared using the SMART complementary DNA library construction kit (Clontech), and sequenced as  $2 \times 100$  paired end runs with Illumina HiSeq 2000 at SciLifeLab (Stockholm, Sweden) or GeneCore (EMBL Genomics Core Facilities). Illumina data were supplemented with publically available RNaseq and genome data (Supplementary Table 1) to generate a final data set including 76 diverse metazoans and 2 choanoflagellate outgroup taxa.

**Data set assembly.** Both novel RNA-seq data and raw Illumina sequences taken from the NCBI Sequence Read Archive were assembled using Trinity<sup>31</sup>. Assembled data were translated using Transdecoder (<http://transdecoder.sf.net>). To determine orthologous genes, we used two methods: a more restrictive and standard approach using HaMStR (Hidden Markov Model based Search for Orthologues using Reciprocity)<sup>32</sup>, as well as an approach designed to generate a broader set of genes for phylogenetic inference, using the software ProteinOrtho<sup>33</sup>. Protocols for gene selection using HaMStR followed refs 34 and 35. Translated unigenes for all taxa were searched against the model organisms core orthologue set of HaMStR using the strict option and *D. melanogaster* as the reference taxon. Sequences shorter than 50 amino acids were deleted, and orthologous groups sampled for fewer than 30 taxa were excluded to reduce missing data. To trim mistranslated ends, if one of the first or last 20 characters of sequences was an X, all characters between that X and the end of the sequence were removed. The orthologous groups were then aligned using MAFFT<sup>36</sup> and trimmed using Aliscore<sup>37</sup> and Alicut (<https://www.zfmk.de/en/research/research-centres-and-groups/utilities>). At this stage, sequences that were greater than 50% gaps and alignments shorter than 100 amino acids were discarded. To remove potentially paralogous genes, we generated single gene trees using FastTree<sup>38</sup> and filtered these using PhyloTreePruner<sup>39</sup>. For 78 taxa, this protocol retained 212 orthologous groups, 44,896 amino acids, with 31% missing data. This protocol was repeated with the 56 taxa with highest percentage of gene coverage, resulting in a data matrix of 336 genes, 81,451 amino acids, and 11% missing data.

To generate the ProteinOrtho data set, *Sterreria* sp. was excluded owing to its small library size. Translated assemblies were filtered to remove mistranslated ends as described above, and only sequences longer than 50 amino acids were retained for clustering in ProteinOrtho. In ProteinOrtho, we used the steps option, the default E-value for BLAST, and minimum coverage of best BLAST alignments of 33%. Resulting clusters were filtered to include only putative orthologous groups containing greater than 40 taxa, then aligned as above with MAFFT. For each alignment a consensus sequence was inferred using the EMBOSS program infoalign<sup>40</sup>. Infoalign's 'change' calculation computes the percentage of positions within each sequence in each alignment that differ from the consensus. Sequences with a 'change' value larger than 75 were deleted, helping to exclude incorrectly aligned sequences. Orthologous groups were then realigned with MAFFT, trimmed with Aliscore and Alicut, and processed as above. After filtering for paralogous groups with PhyloTreePruner, 881 orthologous groups were retained.

Owing to the smaller size of the data set and amount of computational resources required, taxon pruning and signal dissection analyses were performed solely on the primary HaMStR gene set. For taxon exclusion experiments, individual orthologous group alignments were realigned using MAFFT following the removal of selected taxa. TreSpEx<sup>41</sup> was used to assess potential sources of misleading signal, including standard deviation of branch-length heterogeneity (LB) and saturation. Sites showing evidence of saturation and compositional heterogeneity were removed using Block Mapping and Gathering with Entropy (BMGE)<sup>42</sup>, using the 'fast' test of compositional heterogeneity (-s FAST) and retaining gaps (-g 1).

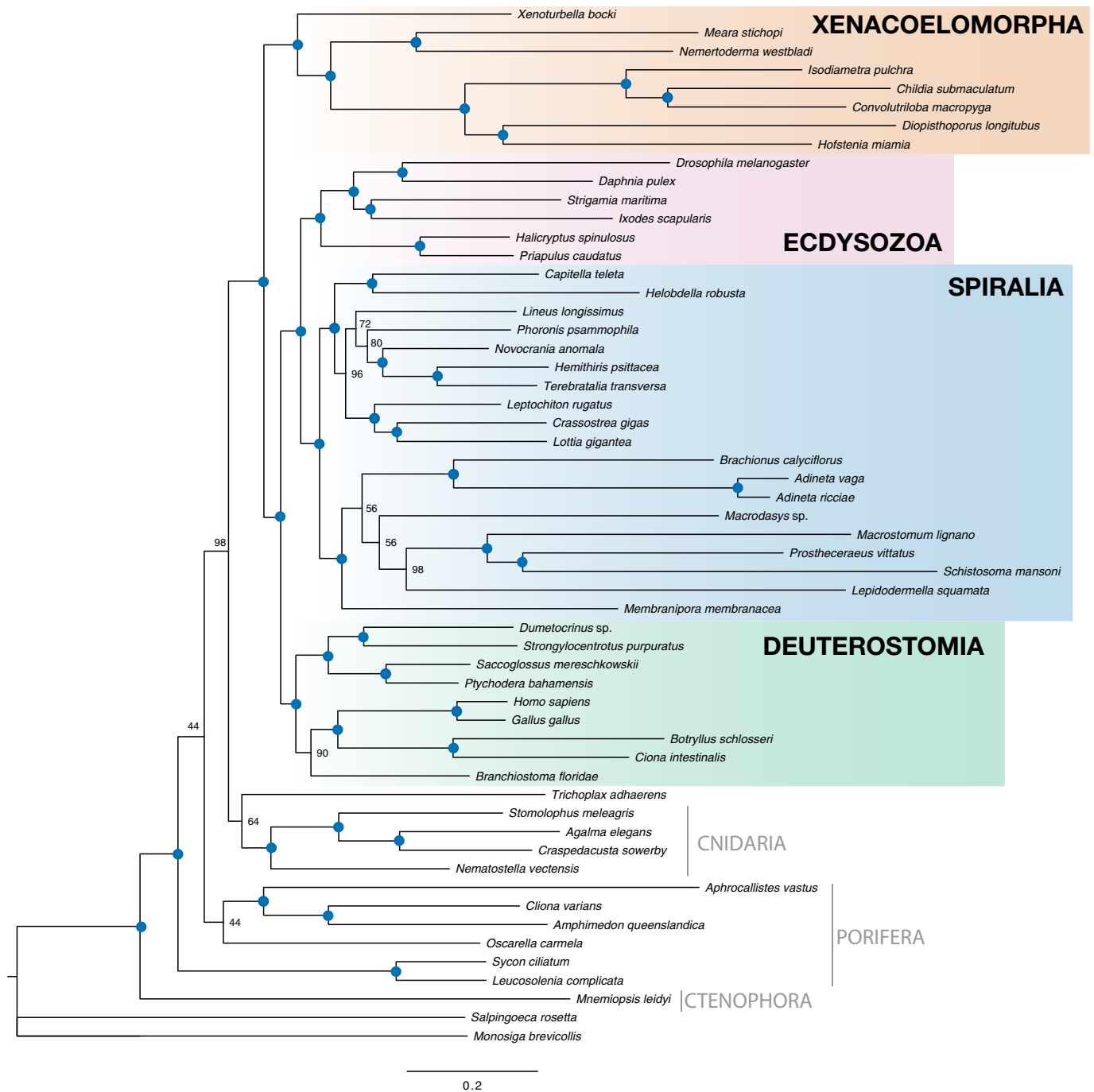
**Phylogenetic analysis.** Maximum likelihood analyses of the complete 212 orthologous group data matrix were performed using RAxML version 8.0.20-mpi<sup>43</sup> under the best-fitting models for each gene partition determined by ProtTest version 3.4 (ref. 44). The best fitting model for all but 3 of the 212 orthologous groups was LG, so further maximum likelihood analyses were performed using the PROTGAMMAILG option. Bootstrapped trees from the 212-gene data set were used to calculate leaf stability indices of each operational taxonomic unit

using the Rogenarok server (<http://www.exelixis-lab.org/>). Bayesian analyses were conducted using PhyloBayes-MPI<sup>45</sup> version 1.5a under the CAT + GTR +  $\Gamma$  model or GTR +  $\Gamma$  with four independent chains per analysis. Analyses ran for >12,000 cycles, until convergence of at least two chains was reached as assessed by maxdiff. Further Bayesian analyses were conducted in MrBayes version 3.2 (ref. 46). For the MrBayes analyses, we back-translated the aligned amino-acid data to nucleotides for first and second codon positions using the universal genetic code. Third codon position data were ignored. When the back translation was ambiguous, we preserved the ambiguity in the nucleotide data. For instance, serine is coded by TC{A, C, G, T} or AG{T, C}, where {...} denotes alternative nucleotides for a single codon site. Thus, for Serine the back translation is {A,T}{C, G}. This is the only back translation that is ambiguous both for the first and for the second codon positions. The back translation for arginine and leucine are also ambiguous but only for the first codon position. All other back translations are unambiguous for both the first and second codon sites. Thus, the back translation of first and second codon sites results in negligible information loss compared with the original nucleotide data.

We analysed the resulting nucleotide data in MrBayes 3.2.6-svn(r1037)<sup>46</sup> using a model with two partitions: one for first codon positions and one for second codon positions. For each partition we employed an independent substitution model, modelling rate variation across sites using a discrete gamma distribution (four categories) with a proportion of invariable sites ('lset rates = invgamma'), and nucleotide substitutions with independent stationary state frequencies and a reversible-jump approach to the partitioning of exchangeability rates ('lset nst = mixed'). We also uncoupled the partition rates ('prset ratepr = variable'). All other settings were left at their defaults. For each analysis, we used four independent runs with four Metropolis-coupled chains each and ran them for 4,000,000 generations, sampling every 500 generations ('mcmc nrun = 4 nch = 4 ngen = 4000000 samplefreq = 500'). The analyses finished with an average standard deviation of split frequencies of 0.033 or less, and a potential scale reduction factor of 1.003 or less. The MrBayes data files and run scripts are provided at the Dryad Digital Repository.

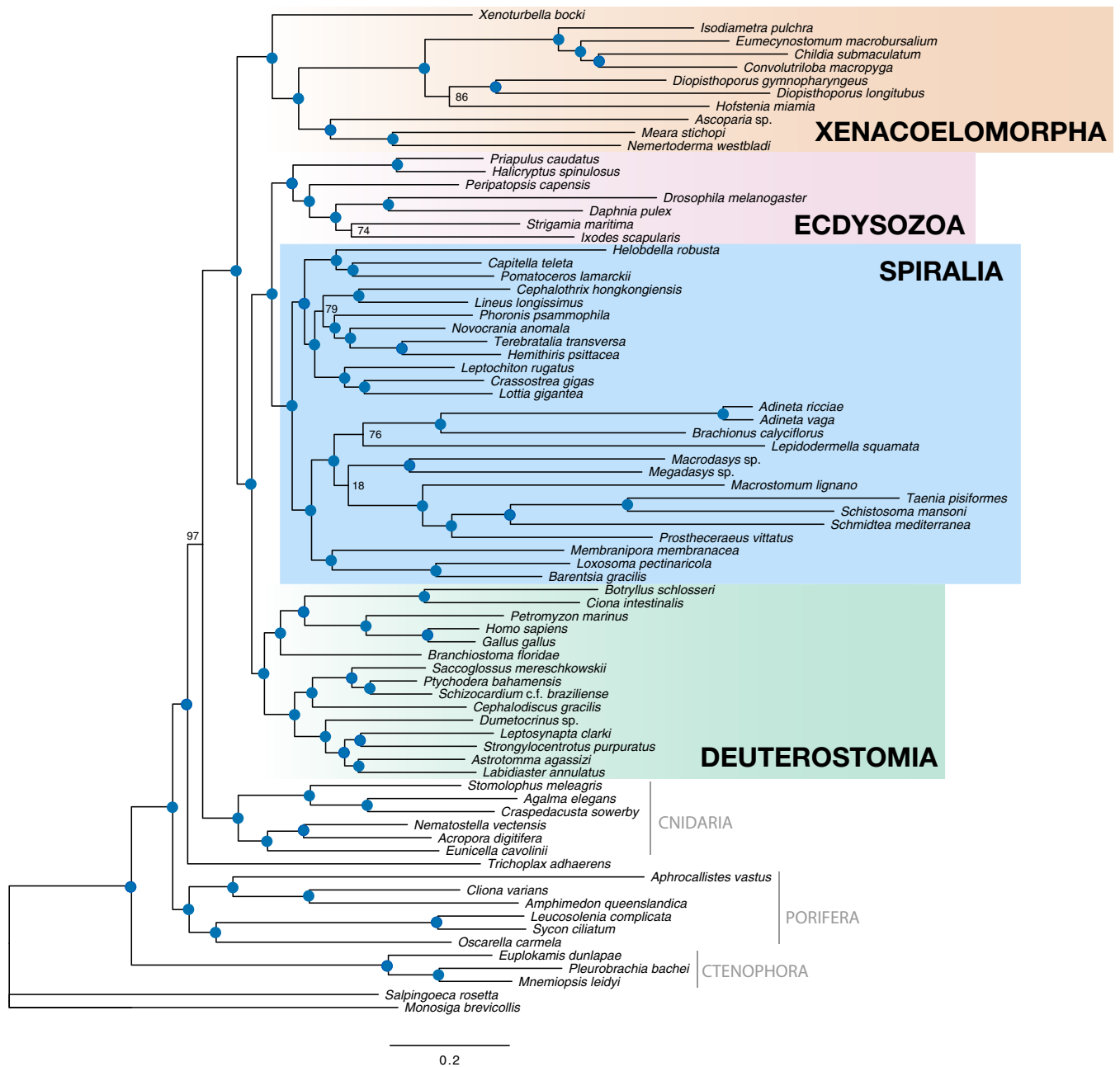
We additionally used ASTRAL<sup>26</sup> to calculate an optimal bootstrapped species tree from individual RAxML gene trees decomposed into quartets.

- Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol.* **29**, 644–652 (2011).
- Ebersberger, I., Strauss, S. & von Haeseler, A. HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC Evol. Biol.* **9**, 157 (2009).
- Lechner, M. et al. Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**, 124 (2011).
- Kocot, K. M. et al. Phylogenomics reveals deep molluscan relationships. *Nature* **477**, 452–456 (2011).
- Cannon, J. T. et al. Phylogenomic resolution of the hemichordate and echinoderm clade. *Curr. Biol.* **24**, 2827–2832 (2014).
- Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
- Misof, B. & Misof, K. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Syst. Biol.* **58**, 21–34 (2009).
- Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
- Kocot, K. M., Citarella, M. R., Moroz, L. L. & Halanach, K. M. PhyloTreePruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evol. Bioinform. Online* **9**, 429–435 (2013).
- Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
- Struck, T. H. TreSpEx-Detection of misleading signal in phylogenetic reconstructions based on tree information. *Evol. Bioinform. Online* **10**, 51–67 (2014).
- Crisuolo, A. & Grihaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
- Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
- Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
- Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615 (2013).
- Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).



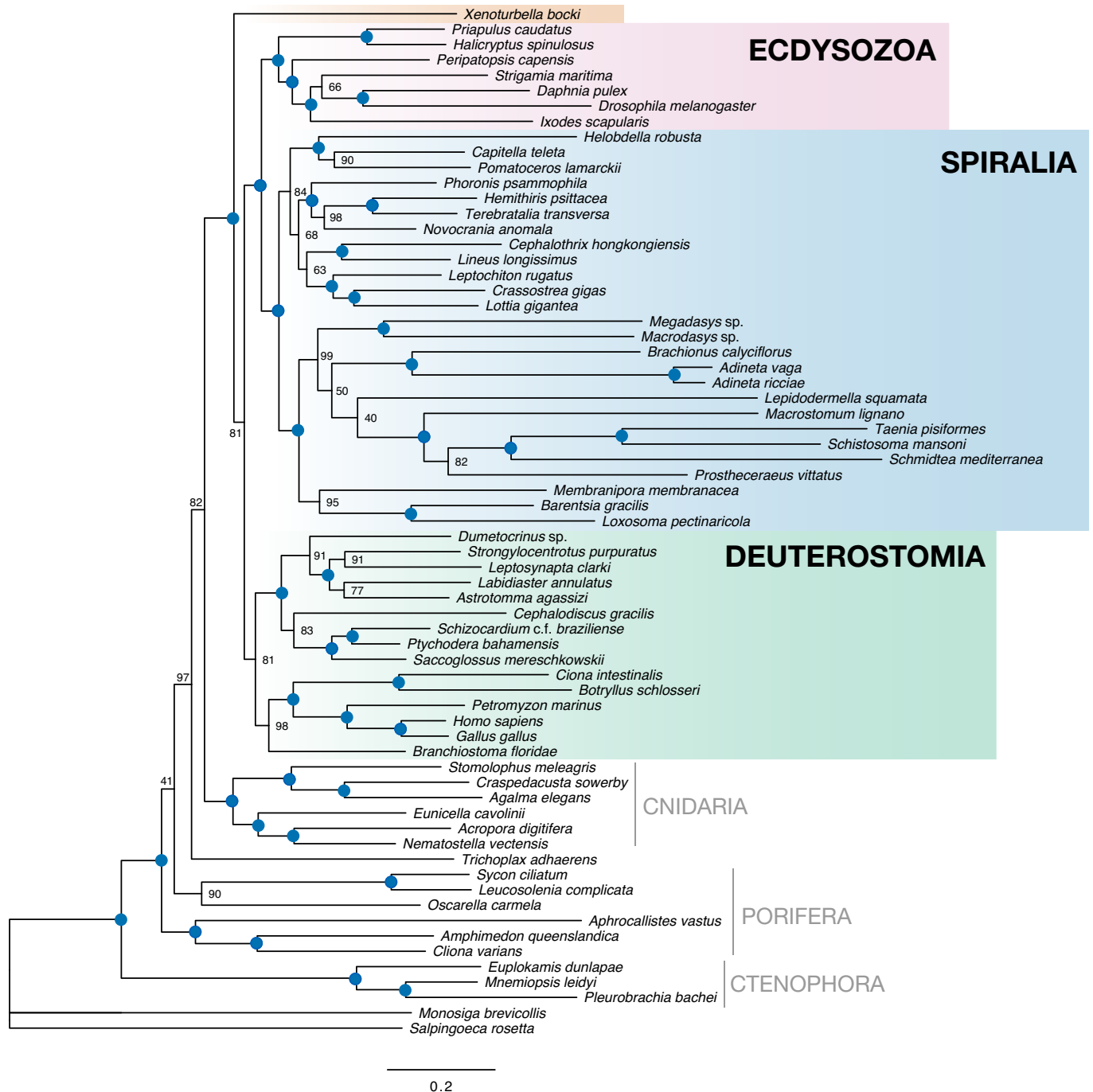
**Extended Data Figure 1 | Maximum likelihood topology of metazoan relationships inferred from 336 genes from the best-sampled 56 taxa.** Maximum likelihood tree is shown as inferred using the LG + I +  $\Gamma$  model

for each gene partition, and 100 bootstrap replicates. Filled blue circles represent 100% bootstrap support. The length of the matrix is 81,451 amino acids and overall matrix completeness is 89%.



**Extended Data Figure 2 | Maximum likelihood topology of metazoan relationships inferred from 881 genes and 77 taxa.** Maximum likelihood tree is shown as inferred using the LG + I +  $\Gamma$  model for each gene

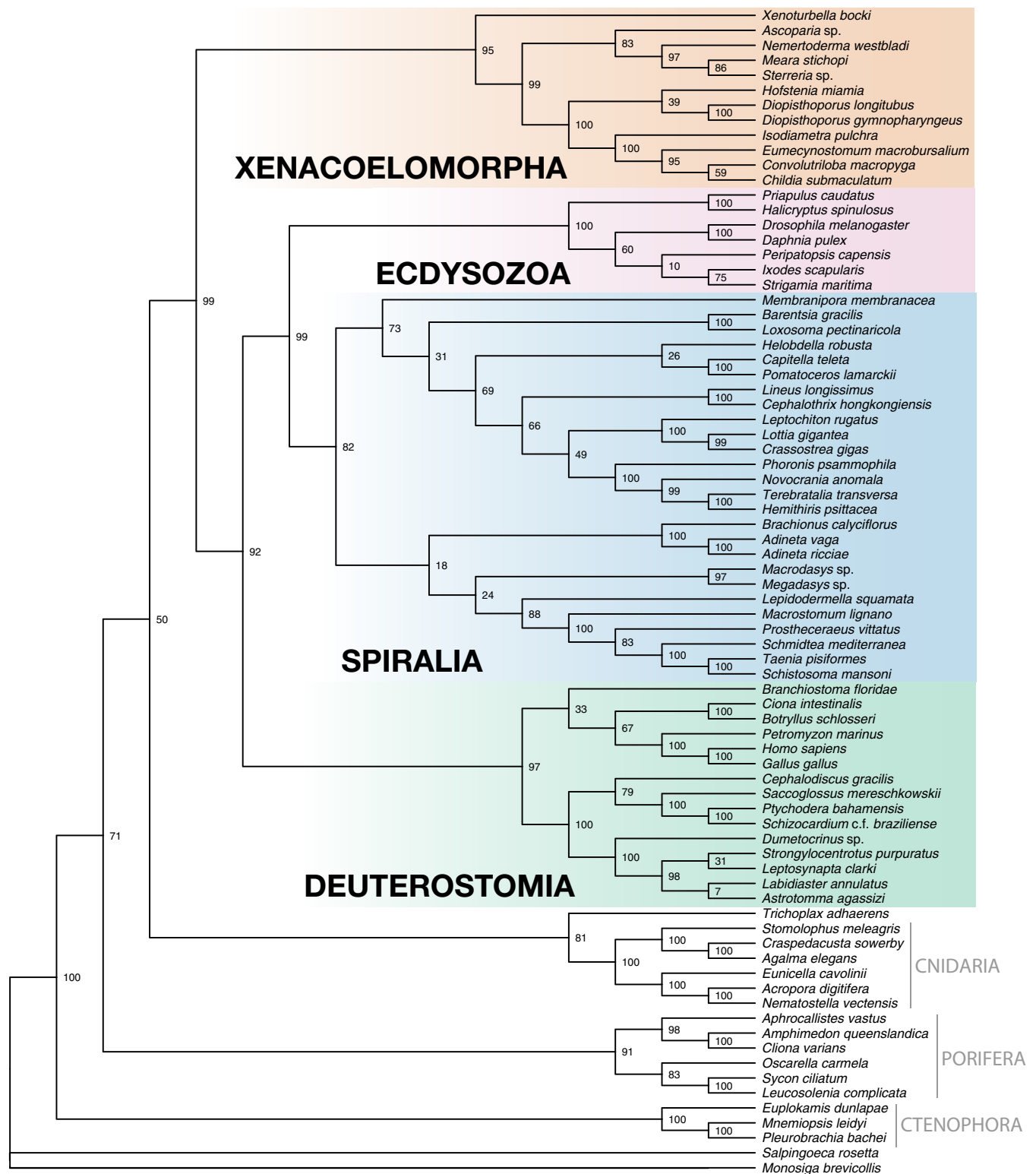
partition, and 100 bootstrap replicates. Filled blue circles represent 100% bootstrap support. The length of the matrix is 337,954 amino acids and overall matrix completeness is 62%.



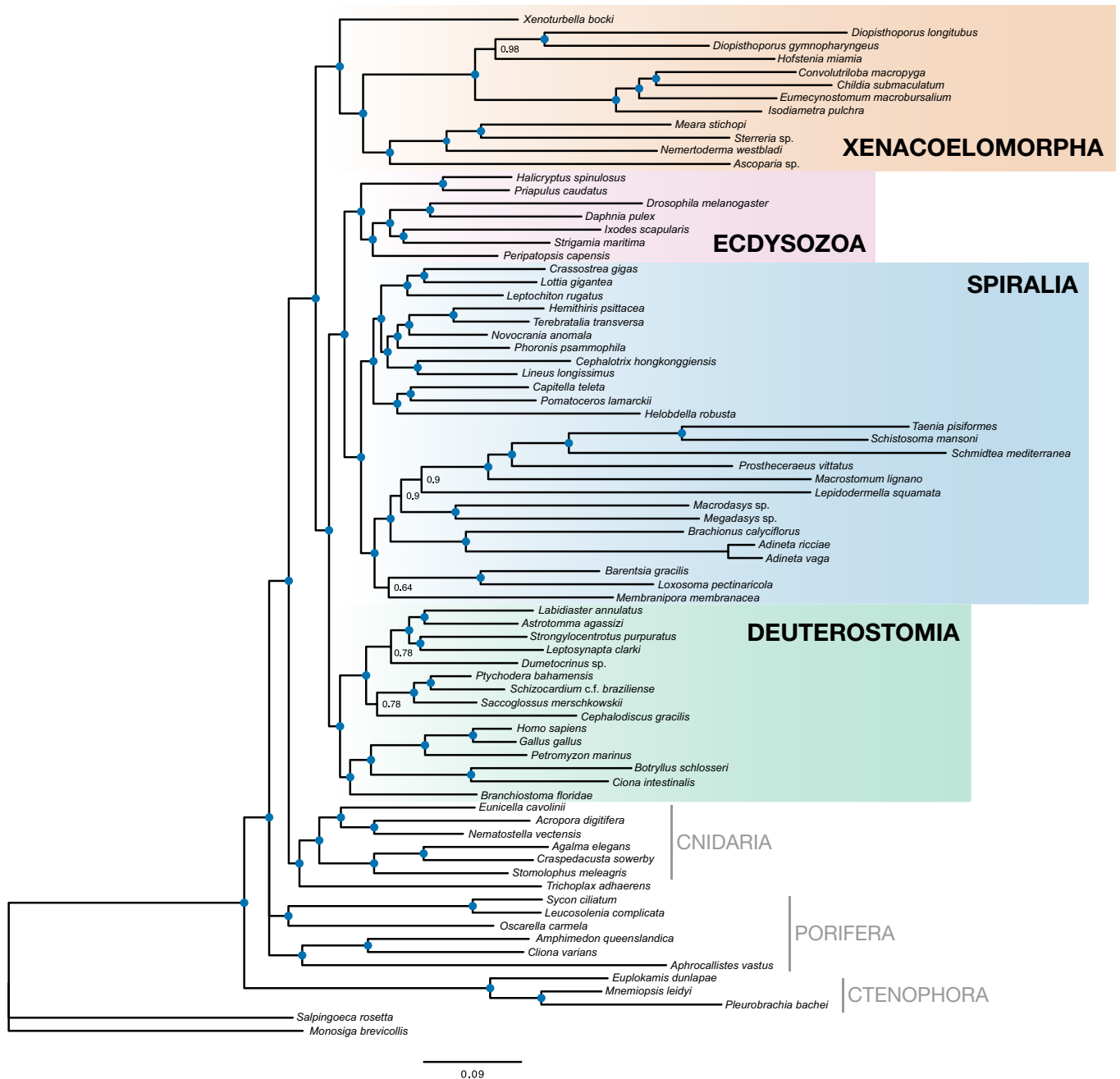
**Extended Data Figure 3 | Maximum likelihood topology of metazoan relationships inferred from 212 genes with Acoelomorpha removed.**  
Maximum likelihood tree is shown as inferred using the LG + I +  $\Gamma$  model

for each gene partition, and 100 bootstrap replicates. Filled blue circles represent 100% bootstrap support. The length of the matrix is 43,942 amino acids and overall matrix completeness is 70%.





**Extended Data Figure 4 |** ASTRAL species tree, constructed from 212 input partial gene trees inferred in RAxML version 8.0.20. Nodal support values reflect the frequency of splits in trees constructed by ASTRAL from 100 bootstrap replicate gene trees.



**Extended Data Figure 5 | Bayesian inference topology of metazoan relationships inferred on the basis of 212 genes and 78 taxa.** Results are shown from MrBayes analyses of four independent Metropolis-coupled

chains run for 4,000,000 generations, with sampling every 500 generations. Amino-acid data were back-translated to nucleotides and analysed under an independent substitution model.

Extended Data Table 1 | Summary of data sets analysed in this study and support for monophyly of major groups

Dataset description	Number of OGs	Number of Taxa	AA positions	% Missing Data	Xenacoelomorpha	Nephrozoa	Bilateria
HaMStR all taxa	212	78	44896	31	100/100/100	99/100/100	100/100/100
HaMStR best coverage taxa	336	56	81451	11	100	100	100
ProteinOrtho	881	77	337954	38	100	100	100
Remove Acoelomorpha	212	67	43942	30	N/A	81	100
Remove <i>Xenoturbella</i>	212	77	43510	31	(Acoelomorpha 100)	70	100
Remove Acoela	212	71	43451	31	100/100	100/1.0	100/100
Remove Nemertodermatida	212	74	45054	30	100	100	100
Remove Ctenophora	212	75	47011	30	100	100	100
Remove Cnidaria	212	72	44990	31	100	100	100
Remove Porifera	212	72	43829	31	100	100	100
Remove Placozoa	212	77	43940	31	100	100	100
Porifera only non-bilaterian Metazoa	212	68	47115	30	100	100	100
Remove non-metazoans	212	76	43764	31	100	100	100
Reduce deuterostomes	210	74	46101	29	100	100	100
Taxa >80% gene occupancy only	212	52	43868	16	100	99	100
Taxa >90% gene occupancy only	212	40	42840	11	100	99	100
Remove taxa with LB score >13	212	59	43247	30	100	98	100
Remove taxa with LB score >30	212	73	44260	30	100	100	100
Genes with best LB scores	106	78	22295	30	100	71	100
Genes with poor LB scores	106	78	22601	32	100	99	100
Genes with lowest saturation	106	78	23414	29	100	95	100
Genes with highest saturation	106	78	21482	34	100	100	100
Only non-ribosomal protein genes	207	78	44715	32	100	100	100
Ribosomal protein genes, LG all partitions	53	78	9010	19	non-monophyletic	non-monophyletic	88
BMGE trimming	Merged	78	33323	34	100	100	100

Bootstrap support values given from RAxML analyses inferred with the LG + I +  $\Gamma$  model from 100 rapid bootstrap replicates. Bayesian posterior probabilities are listed from MrBayes analyses inferred under an independent substitution model using a back-translated nucleotide data set derived from our amino-acid alignment, and PhyloBayes analyses under the CAT + GTR +  $\Gamma$  model.

# New deep-sea species of *Xenoturbella* and the position of Xenacoelomorpha

Greg W. Rouse<sup>1</sup>, Nerida G. Wilson<sup>1,2,3</sup>, Jose I. Carvajal<sup>1</sup> & Robert C. Vrijenhoek<sup>4</sup>

The discovery of four new *Xenoturbella* species from deep waters of the eastern Pacific Ocean is reported here. The genus and two nominal species were described from the west coast of Sweden<sup>1,2</sup>, but their taxonomic placement remains unstable<sup>3,4</sup>. Limited evidence placed *Xenoturbella* with molluscs<sup>5,6</sup>, but the tissues can be contaminated with prey<sup>7,8</sup>. They were then considered deuterostomes<sup>9–13</sup>. Further taxon sampling and analysis have grouped *Xenoturbella* with acoelomorphs (=Xenacoelomorpha) as sister to all other Bilateria (=Nephrozoa)<sup>14,15</sup>, or placed Xenacoelomorpha inside Deuterostomia with Ambulacraria (Hemichordata + Echinodermata)<sup>16</sup>. Here we describe four new species of *Xenoturbella* and reassess those hypotheses. A large species (>20 cm long) was found at cold-water hydrocarbon seeps at 2,890 m depth in Monterey Canyon and at 1,722 m in the Gulf of California (Mexico). A second large species (~10 cm long) also occurred at 1,722 m in the Gulf of California. The third large species (~15 cm long) was found at ~3,700 m depth near a newly discovered carbonate-hosted hydrothermal vent in the Gulf of California. Finally, a small species (~2.5 cm long), found near a whale carcass at 631 m depth in Monterey Submarine Canyon (California), resembles the two nominal species from Sweden. Analysis of whole mitochondrial genomes places the three larger species as a sister clade to the smaller Atlantic and Pacific species. Phylogenomic analyses of transcriptomic sequences support placement of Xenacoelomorpha as sister to Nephrozoa or Protostomia.

*Xenoturbellida* Bourlat *et al.*, 2006

Genus *Xenoturbella* Westblad, 1949

*Xenoturbella monstrosa* sp. nov.

**Etymology.** Latin for extraordinary size.

**Material examined.** Holotype, Scripps Institution of Oceanography Benthic Invertebrate Collection SIO-BIC BI1037, sex unknown (Extended Data Fig. 1c). Paratype SIO-BIC BI1038, from type locality, sex unknown. Paratype SIO-BIC BI1039, female, from Mexican locality (Fig. 1b and Extended Data Fig. 1d–h).

**Locality.** Monterey Submarine Canyon, California, 36° 36.8' N, 122° 26.0' W, ~2,890 m depth, vesicomyid clam field (Extended Data Fig. 1a, b). Collected via slurp system, research vessel (R/V) *Western Flyer*/remote operated vehicle (ROV) *Tiburon* dive 610, 25 October 2004. Also from Guaymas Transform Fault, Gulf of California, Mexico, 27° 34.659' N, 111° 27.036' W, 1,722 m depth, near a cold-water methane seep, 'Pinkies Vent'. Collected via slurp, R/V *Western Flyer*/ROV *Doc Ricketts* dive 385, 13 April 2012.

**Description.** Body ~20 cm long in life, purple or pale pink. Dorsal body wall with two deep, longitudinal furrows. Ring furrow and side furrow found. Body in front of ring furrow rounded; posterior tapers gradually. Mouth oval-shaped (when relaxed), lying ventrally midway between ring furrow and anterior end. Epidermal network over two-thirds of ventral surface. Body wall with gametes dorsally and ventrally. Exogenous DNA present in tissue from co-occurring vesicomyid

bivalves *Archivesica diagonalis* (in holotype) or *Calyptogena pacifica* (in Mexican paratype).

*Xenoturbella churro* sp. nov.

**Etymology.** Noun used in apposition. Resembles fried-dough pastry called churro.

**Material examined.** Holotype (SIO-BIC BI1040), female (Fig. 1b and Extended Data Fig. 2a–e).

**Locality.** Guaymas Transform Fault, Gulf of California, Mexico, 27° 34.659' N, 111° 27.036' W, 1,722 m depth, near 'Pinkies Vent'. Collected via slurp, R/V *Western Flyer*/ROV *Doc Ricketts* dive 385, 13 April 2012.

**Description.** Body 10 cm long in life, uniformly orange/pink. Dorsal body wall with four deep, longitudinal furrows. Ring furrow and side furrow present. Anterior end rounded, posterior tapers sharply. Mouth oval (when relaxed), lying ventrally midway between ring furrow and anterior end. Epidermal network over two-thirds of ventral surface. Body wall with gametes dorsally and ventrally. Exogenous DNA from co-occurring vesicomyid bivalve *C. pacifica* present in tissue.

*Xenoturbella profunda* sp. nov.

**Etymology.** Latin for deep.

**Material examined.** Holotype SIO-BIC BI1041, male (Extended Data Fig. 3b, c, g, h). Paratypes (SIO-BIC BI1042–46, UNAM ICML-EMU-11010), a female and five males or juveniles (Fig. 1c and Extended Data Fig. 3a, d–f), from type locality.

**Locality.** Pescadero Basin, Mexico, 23° 57.23' N, 108° 51.73' W, 3,700 m depth on sediment near hydrothermal vent. Collected via slurp, R/V *Western Flyer*/ROV *Doc Ricketts* over several dives in April 2015.

**Description.** Body to 15 cm in life, uniformly pale pink, with epidermis, circular and longitudinal muscles, parenchyma and gastrodermis. Dorsal body with a pair of deep, longitudinal furrows. Ring furrow and side furrow found. Anterior end rounded, posterior tapering gradually. Mouth oval (when relaxed), ventral, anterior to ring furrow. Epidermal network over two-thirds of ventral surface. Gametes dorsally and ventrally. Oocytes reach 450 µm in diameter, sperm with spherical heads. Exogenous DNA of vesicomyid bivalve *A. gigas* present in tissue.

*Xenoturbella hollandorum* sp. nov.

**Etymology.** Named for Linda and Nicholas Holland for their contributions to biology.

**Material examined.** Holotype SIO-BIC BI1036, sex unknown (Fig. 1d and Extended Data Fig. 4a–d).

**Locality.** Monterey Submarine Canyon, California, 36° 48.132' N, 121° 59.647' W; ~631 m depth, on sediment adjacent to bones of a grey whale. Collected via push core, R/V *Western Flyer*/ROV *Tiburon* dive 1,160, 18 December 2007.

**Description.** Body short (2.5 cm) and uniformly bright pink. Dorsal body with a pair of longitudinal furrows. Ring furrow and side furrow found. Ventral mouth present, diamond-shaped (when relaxed), just anterior to ring furrow. Epidermal network over ventral surface anterior to ring furrow. Exogenous cytochrome-*c*-oxidase subunit I (*COI*) sequences of bivalves not detected.

<sup>1</sup>Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92037, USA. <sup>2</sup>Western Australian Museum, Locked Bag 49, Welshpool DC, Western Australia 6986, Australia. <sup>3</sup>School of Animal Biology, University of Western Australia, Crawley, Western Australia 6009, Australia. <sup>4</sup>Monterey Bay Aquarium and Research Institute, Moss Landing, California 95039, USA.





**Figure 1 | New species of *Xenoturbella*.** **a**, *X. monstrosa* (arrows) in a field of *A. diagonalis* (Ad) and *Ectenagena elongata* (Ee) clams at 2,890 m depth in Monterey Bay, California. Scale estimated from the average size of *A. diagonalis*<sup>27</sup>. **b**, *X. monstrosa* (arrow at upper left) and *X. churro* (asterisk at lower right) at 1,722 m depth near a methane seep in the Guaymas Basin, Mexico. Red spots (lasers) are 29 cm apart. Shell fragments in sediment,

but no living clams observed in vicinity. Both specimens contained *C. pacifica* DNA in their tissues. **c**, *X. profunda* female in an *A. gigas* (Ag) clam field near a carbonate-hosted hydrothermal vent at ~3,700 m depth in Pescadero Basin, Mexico. **d**, *X. hollandorum* on sediment adjacent to whale bones (b). Abbreviations: a, anterior; rf, ring furrow; h, hemichordate; o, bone-eating *Osedax* worms; p, polynoid scaleworm; s, shrimp.

While we have markedly increased the diversity of the group, the morphological similarities of the new species leads us to keep them in *Xenoturbella*<sup>1</sup>. Live specimens of all four new species exhibited an epidermal branching network ventrally that has not been previously recognized. This network was inconspicuous in *X. hollandorum* and consequently may have been overlooked previously in *X. bocki*. *X. profunda* is gonochoric, as opposed to the hermaphroditism reported for *X. bocki*, which also has much smaller eggs<sup>1,4</sup>. *X. monstrosa* was seen in abundance once (Fig. 1a, Extended Data Fig. 1a, b and Supplementary Video 1) in a vesicomyid clam field (mostly *Ectenagena extenta*, some *A. diagonalis*) in Monterey Canyon, but repeated visits to this site failed to reveal more specimens. The *X. monstrosa* collected from Monterey were in a patch of mainly *A. diagonalis* that were dying or spawning, producing large amounts of mucus (Supplementary Video 1). The Gulf of California specimen extends the range of *X. monstrosa* to >2,500 km and depth from ~1,700 to ~3,000 m. The single specimen of *X. churro* was found only 30 cm from the Mexican specimen of *X. monstrosa* (Fig. 1b and Supplementary Video 2). *Xenoturbella profunda* lives deepest of the known species, and all seven collected specimens were observed on sediments (showing bacterial mats and various fauna, including vesicomyids) near a hydrothermal vent (Supplementary Video 3).

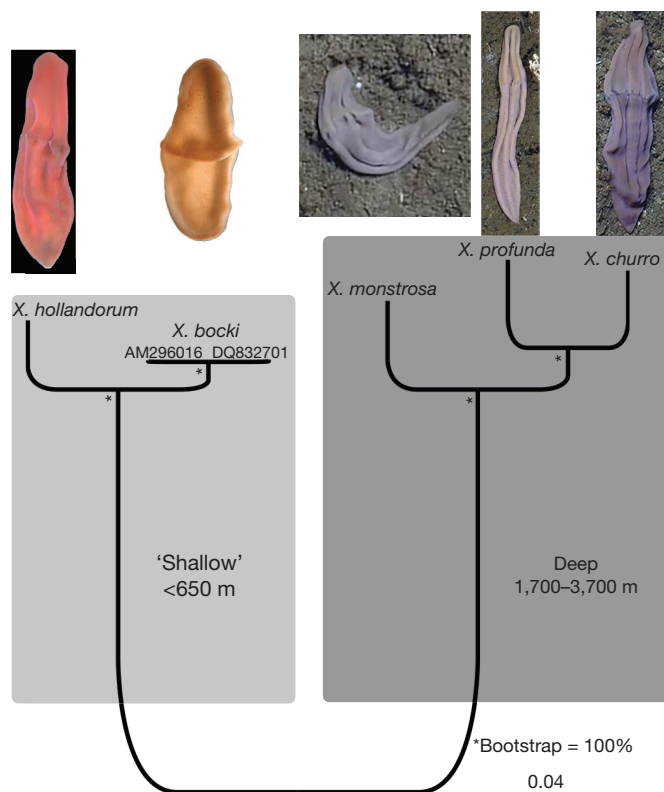
Mitochondrial *COI* was sequenced for all specimens examined and compared with other *Xenoturbella*. The nominal species from Sweden, *X. bocki* and *X. westbladi*, comprised a single haplotype network with several shared sequences (Extended Data Fig. 5a), which strongly suggests it is a single species, although this has previously been dismissed<sup>17</sup>. Here we treat *X. westbladi* as a junior synonym of *X. bocki*. The two specimens of *X. monstrosa* from California differed by one base pair

(bp) and from the Mexican specimen by up to 7 bp (Extended Data Fig. 5b). Seven specimens of *X. profunda* exhibited four haplotypes, differing by one to three bp (Extended Data Fig. 5c).

Whole mitochondrial genomes of the four new species (Extended Data Fig. 6) exhibited the same gene order as *X. bocki*<sup>9,18</sup>. Phylogenetic analysis of these genomes placed the smaller, shallower-dwelling species, *X. hollandorum* and *X. bocki*, as sister taxa differing at ≥6.4% of the nucleotide sites and 59 amino acids. The distinction of *X. hollandorum* as a new species is also supported by nuclear sequence data (Supplementary Table 1). Among all other *Xenoturbella* species, mitochondrial genomes differed by 10–20% (Supplementary Data Table 1), with the deep-water species forming a clade with respect to *X. hollandorum* and *X. bocki* (Fig. 2). Phylogenetic comparison of mitochondrial proteins placed Xenacoelomorpha with deuterostomes, as previously reported<sup>9,16,18</sup> (Extended Data Fig. 7), although poorly supported. Notably, the PhyloBayes<sup>19</sup> analysis recovered Xenacoelomorpha as sister group to Chordata, also with poor support, and we question the utility of such data at deep phylogenetic scales.

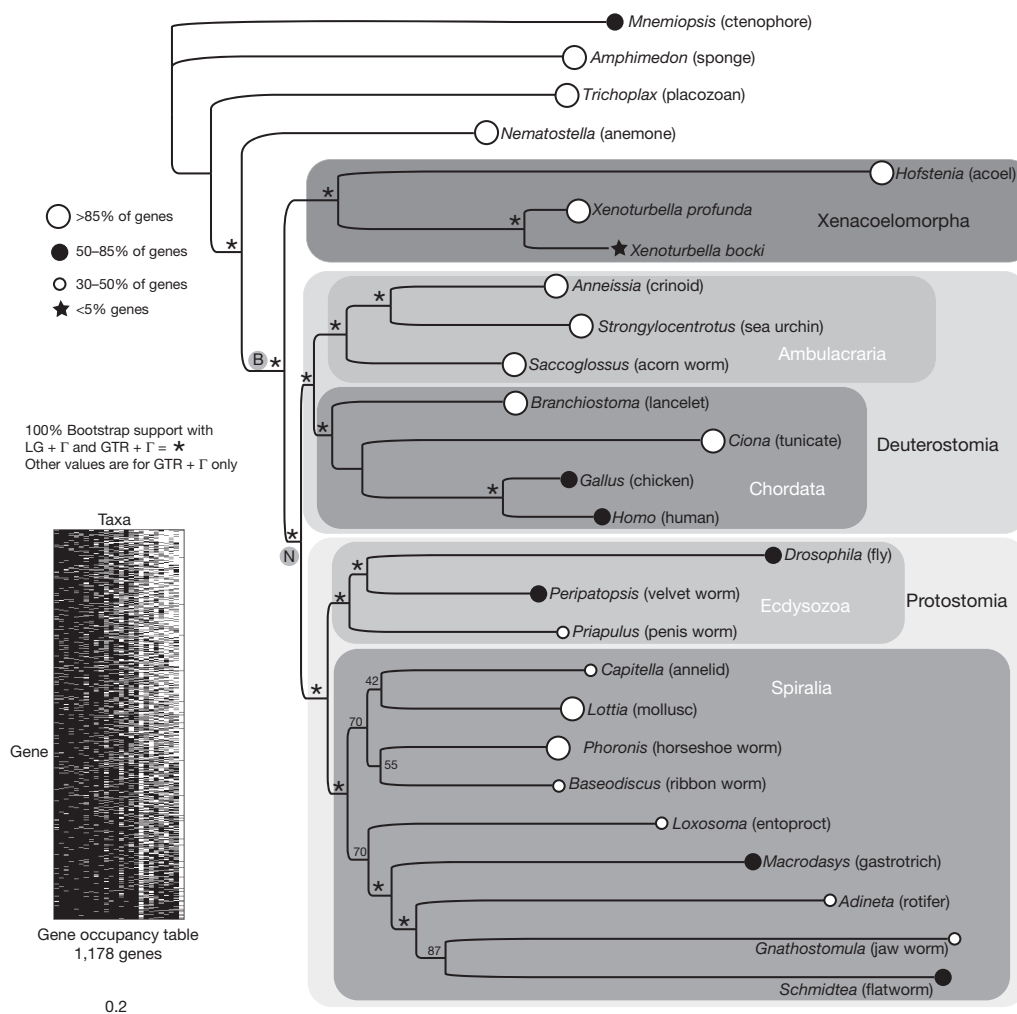
Phylogenomic analyses of *Xenoturbella* compared our transcriptomic data from *X. profunda* with limited transcriptomic (expressed sequence tag) data from *X. bocki*<sup>11</sup>, and genomic and transcriptomic data from other animals. The main data set included 26 terminals and 1,178 genes with 70% average gene occupancy (Supplementary Table 5). Maximum likelihood analyses using RAXML<sup>20</sup> with either the WAG, LG or GTR amino-acid substitution models and gamma (Γ) correction for rate heterogeneity produced identical tree topologies and similar support values, grouping *Xenoturbella* with the acoel *Hofstenia* (=Xenacoelomorpha) as sister to Nephrozoa





**Figure 2 | Phylogeny of *Xenoturbella* based on mitochondrial genomic DNA sequences (15,532 base-pair alignment) using RAxML and GTR +  $\Gamma$ . Support assessed via 100 pseudo-replicates of bootstrapping. Scale bar, substitutions per nucleotide position. All nodes were highly supported as indicated by bootstrap support of 100% (asterisk). Rooting based on the topology recovered from the analysis of mitogenomes of Metazoa (Extended Data Fig. 7). Scale bar, substitutions per nucleotide position.**

(deuterostomes + protostomes; Fig. 3). All key nodes were well supported. Alternative hypotheses assessed using SOWH<sup>21</sup> constraining Xenacoelomorpha as sister group to Ambulacraria ( $n = 100$ ,  $\Delta$ -likelihood = 1,313.9,  $P = 0.027$ ), or to Deuterostomia ( $n = 100$ ,  $\Delta$ -likelihood = 1,042.07,  $P = 0.031$ ), were rejected as significantly worse. A species tree<sup>22</sup> generated from the consensus of 393 individual gene-trees (average occupancy = 80%) (Extended Data Fig. 8) placed Xenacoelomorpha as sister to Nephrozoa, although there was poor support for the latter clade. We also used PhyloBayes under CAT + GTR +  $\Gamma$ , as recommended in a previous assessment of the position of *Xenoturbella*<sup>16</sup>. The result placed Xenacoelomorpha as sister to protostomes with moderate support, a placement not previously recovered (Extended Data Fig. 9). To explore if elevated substitution rates in some loci could be influencing our analyses<sup>23</sup>, we removed the ribosomal proteins (51 loci) from the 1,178-locus data set and then split this into slowest- and fastest-evolving matrices. Maximum likelihood analyses revealed no major differences between the two data sets or incongruence with the complete data set (Extended Data Fig. 10).



**Figure 3 | Maximum likelihood phylogeny of animals based on 1,178 genes (394,818 amino acids) with 70% average gene occupancy. Within Bilateria (B), Xenacoelomorpha is sister group to Nephrozoa (N), which includes Protostomia and Deuterostomia. Key nodes were highly supported, as indicated by bootstrap support of 100% (asterisks). Data were analysed using RAxML with GTR, LG and WAG models, +  $\Gamma$ . Support values of less than 100 are shown for GTR +  $\Gamma$ ; the same nodes had higher support with LG +  $\Gamma$  and WAG +  $\Gamma$ . Scale bar, substitutions per amino-acid position.**

We sampled four new *Xenoturbella* species from cold-water hydrocarbon seeps, a hydrothermal vent, and from the vicinity of a whale-fall, all environments that host dense communities of animals. These discoveries highlight the possibility of further *Xenoturbella* discoveries in deep-sea environments that sustain their likely food source, bivalve molluscs. Although the placement of *Xenoturbella* has been a challenge in metazoan phylogenetics, the present application of transcriptomic data resulted in much larger sampling of genes, which has helped to resolve deep phylogenetic relationships<sup>24</sup>. Our analyses showed no support for the hypothesis that Xenacoelomorpha lies inside deuterostomes<sup>7,9,16</sup>, a result also corroborated by another phylogenomic study on the position of Xenacoelomorpha<sup>25</sup>. The placement of Xenacoelomorpha as sister group to Nephrozoa (Fig. 3 and Extended Data Figs 8 and 10), or to Protostomia (Extended Data Fig. 9), suggests the suite of features previously proposed to be character losses in *Xenoturbella*<sup>13</sup> appear instead to be plesiomorphic absences.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 19 September; accepted 15 December 2015.**

- Westblad, E. *Xenoturbella bocki* n.g., n.sp. a peculiar, primitive turbellarian type. *Ark. Zool.* **1**, 11–29 (1949).
- Israelsson, O. New light on the enigmatic *Xenoturbella* (phylum uncertain): ontogeny and phylogeny. *Proc. R. Soc. Lond. B* **266**, 835–841 (1999).
- Reisinger, E. Was ist *Xenoturbella*? *Z. Wiss. Zool.* **164**, 188–198 (1960).
- Nakano, H. What is *Xenoturbella*? *Zool. Lett.* **1**, 22 (2015).
- Norén, M. & Jondelius, U. *Xenoturbella*'s molluscan relatives... *Nature* **390**, 31–32 (1997).
- Israelsson, O. ...and molluscan embryogenesis. *Nature* **390**, 32 (1997).
- Bourlat, S. J., Nielsen, C., Lockyer, A. E., Littlewood, D. T. J. & Telford, M. J. *Xenoturbella* is a deuterostome that eats molluscs. *Nature* **424**, 925–928 (2003).
- Bourlat, S. J. *et al.* Feeding ecology of *Xenoturbella bocki* (phylum Xenoturbellida) revealed by genetic barcoding. *Mol. Ecol. Resour.* **8**, 18–22 (2008).
- Bourlat, S. J. *et al.* Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85–88 (2006).
- Bourlat, S. J., Nielsen, C., Economou, A. D. & Telford, M. J. Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. *Mol. Phylogenet. Evol.* **49**, 23–31 (2008).
- Dunn, C. W. *et al.* Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–749 (2008).
- Bourlat, S. J., Rota-Stabelli, O., Lanfear, R. & Telford, M. J. The mitochondrial genome structure of *Xenoturbella bocki* (phylum Xenoturbellida) is ancestral within the deuterostomes. *BMC Evol. Biol.* **9**, 107 (2009).
- Telford, M. J., Budd, G. E. & Philippe, H. Phylogenomic insights into animal evolution. *Curr. Biol.* **25**, R876–R887 (2015).
- Hejnol, A. *et al.* Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc. R. Soc. B* **276**, 4261–4270 (2009).
- Ryan, J. F. *et al.*; NISC Comparative Sequencing Program. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* **342**, 1242592 (2013).
- Philippe, H. *et al.* Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. *Nature* **470**, 255–258 (2011).
- Israelsson, O. & Budd, G. E. Eggs and embryos in *Xenoturbella* (phylum uncertain) are not ingested prey. *Dev. Genes Evol.* **215**, 358–363 (2005).
- Perseke, M. *et al.* The mitochondrial DNA of *Xenoturbella bocki*: genomic architecture and phylogenetic analysis. *Theory Biosci.* **126**, 35–42 (2007).
- Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).
- Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
- Church, S. H., Ryan, J. F. & Dunn, C. W. Automation and evaluation of the SOWH Test with SOWHAT. *Syst. Biol.* **64**, 1048–1058 (2015).
- Mirarab, S. & Warnow, T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**, i44–i52 (2015).
- Whelan, N. V., Kocot, K. M., Moroz, L. L. & Halanych, K. M. Error, signal, and the placement of Ctenophora sister to all other animals. *Proc. Natl Acad. Sci. USA* **112**, 5773–5778 (2015).
- Dunn, C. W., Giribet, G., Edgecombe, G. D. & Hejnol, A. Animal phylogeny and its evolutionary implications. *Annu. Rev. Ecol. Evol. Syst.* **45**, 371–395 (2014).
- Cannon, J. T. *et al.* Xenacoelomorpha is the sister group to Nephrozoa. *Nature* <http://dx.doi.org/10.1038/nature16520> (this issue).
- Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Proc. Gateway Computing Environments Workshop (GCE)*, 14 November 2010, New Orleans, LA, 1–8 (IEEE, 2010).
- Barry, J. P. & Kochevar, R. E. *Calyptogena diagonalis*, a new vesicomyid bivalve from subduction zone cold seeps in the Eastern North Pacific. *Veliger* **42**, 117–123 (1999).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the crew of the R/V *Western Flyer* and pilots of the ROVs *Tiburion* and *Doc Ricketts* for their skill and patience during hunts for these ‘purple socks’. We also thank S. Johnson for verifying bivalve sequences obtained from *Xenoturbella* and L. Lundsten for hunting through many video files for imagery. We acknowledge the Cyberinfrastructure for Phylogenetic Research (CIPRES) Science Gateway<sup>26</sup> for computing resources, and thank M. Miller for additional resources, S. Mirarab for discussions on species tree methods and N. Holland for comments on the manuscript. This work was supported by the David and Lucile Packard Foundation via the Monterey Bay Aquarium Research Institute, Scripps Institution of Oceanography and the National Science Foundation Assembling the Tree of Life program (DEB1036368 to G.W.R.).

**Author Contributions** G.W.R., N.G.W. and R.C.V. collected the specimens. N.G.W. and J.I.C. generated and assembled mitochondrial data for the new *Xenoturbella* species. J.I.C. generated the *Xenoturbella* Illumina transcriptome and assembled the data. G.W.R. and J.I.C. performed phylogenetic analyses of mitochondrial genomes and transcriptome data. G.W.R. analysed the morphology of *Xenoturbella* spp. for the taxonomic descriptions. G.W.R. drafted the paper with J.I.C., N.G.W. and R.C.V. All authors commented on the manuscript.

**Author Information** Sequence data have been deposited in GenBank; accession numbers can be found in Supplementary Tables 3–5. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.W.R. ([grouse@ucsd.edu](mailto:grouse@ucsd.edu)).

## METHODS

No statistical methods were used to predetermine sample size.

**Collecting details.** Once on board the ship, specimens were maintained in chilled seawater in a 4°C cold-room. *Xenoturbella* was photographed under a Leica MZ8 stereomicroscope. Larger animals were photographed with a macro lens. Animals were then relaxed in 7% MgCl in freshwater before further photography. Pieces of tissue were then taken for molecular sequencing and frozen or placed into RNAlater or 95% ethanol. Other pieces of tissue were fixed in 4% paraformaldehyde in 0.2 M sodium phosphate buffer overnight before rinsing and being placed in buffer with sodium azide or processed further and kept in 70% ethanol. One specimen of *X. monstrosa* was fixed and preserved in 95% ethanol and the other in 10% formalin in seawater. All other specimens were fixed in 4% paraformaldehyde in buffer and preserved in 70% ethanol. All specimens are lodged at the Scripps Institution of Oceanography Benthic Invertebrate Collection (SIO-BIC), La Jolla, California, except for a paratype of *X. profunda*, which is lodged at the Instituto de Ciencias del Mar y Limnología (UNAM), Mazatlán, Mexico, and catalogued as ICMLE-EMU11010. Taxonomic acts are registered on Zoobank (<http://zoobank.org/>) as urn:lsid:zoobank.org:pub:135BCE9-A1E4-4926-B346-3711A6485690.

**Amplification and sequencing of mitochondrial and nuclear genes and mitochondrial genomes.** We extracted DNA from *Xenoturbella* individuals using a DNeasy Blood & Tissue Kit following the manufacturer's specifications (Qiagen). *COI* for all *Xenoturbella* individuals and their bivalve prey was amplified with the primers listed in Supplementary Table 2. Histone H3 (HH3) was also amplified with the primers listed in Supplementary Table 2. NCBI accession numbers are listed in Supplementary Table 3. To acquire *Xenoturbella* mitochondrial genomes, we initially amplified portions of three genes: 16S rRNA (16S), cytochrome B (*CytB*) and cytochrome-c-oxidase subunit III (*COIII*). Primer sequences and thermocycling conditions are provided in Supplementary Table 2. All amplifications were done using illustra PuReTaq Ready-To-Go PCR Beads (GE Life Sciences) following the manufacturer's protocol. PCR products were cleaned using USB ExoSAP-IT, sequenced by Eurofins MWG Operon. Geneious R7 (ref. 28) was used to inspect and trim sequences.

We then used the 16S, *CytB* and *COIII* sequences and the Primer3 (ref. 29) algorithm in Geneious to design *Xenoturbella*-specific primers for long PCR amplifications. The mitochondrial genomes of *X. monstrosa*, *X. hollandorum* and *X. churro* were amplified in two overlapping fragments that were each 8,000 bp long using primers listed in Supplementary Table 2. All long PCR products were amplified using Platinum Taq DNA Polymerase High Fidelity (Invitrogen) following the manufacturer's specifications. The PCR products were visualized on 0.9% agarose gels run at 80 V for 90 min. PCR products were cleaned using USB ExoSAP-IT or GelElute Extraction kit (5 Prime). The *X. hollandorum* mitochondrial genome was outsourced to Engencore (Selah Genetics) for sequencing and assembly with the Roche 454 platform and Newbler version 2.3. The *X. monstrosa* and *X. churro* mitochondrial genomes were sequenced by Macrogen using Illumina HiSeq2000. The reads were assembled *de novo* after low-quality reads and adaptor sequences were removed using Geneious. The *X. profunda* mitochondrial genome was obtained from the transcriptome data. *X. profunda* and *X. churro* had fragments of the mitochondrial genome that were not sequenced. These missing 'fill-in' fragments were recovered via direct sequencing using primers listed in Supplementary Table 2 (refs 12, 30–32).

**RNA extraction and transcriptome sequencing.** *X. profunda* tissue was finely chopped and placed in RNAlater (Ambion) for 24 h at 4°C and subsequently at –80°C for long-term storage. RNA was extracted using Direct-zol RNA Kits (Zymo Research) following the manufacturer's protocol. The optional DNase I step was performed during the RNA extraction to remove residual genomic DNA. RNA extractions were then purified using RNA Clean & Concentrator kits (Zymo Research) following the manufacturer's protocol. Libraries were prepared from the purified mRNA using KAPA Stranded mRNA-Seq Kits (Kapa Biosystems) following the manufacturer's protocol. Libraries were sequenced (150 bp paired-end) using a 300 cycle Miseq kit (Illumina) by the IGM Genomics Center (University of California, San Diego). 9,941,202 read pairs were generated from this run, which were then assembled using Trinity into 31,374 transcripts after low-quality, adaptor-contaminated or ribosomal rRNA reads were removed.

**Mitochondrial data analysis.** We used TCS<sup>33</sup> to construct *COI* haplotype networks for the *Xenoturbella* species with multiple sequences. The mitochondrial phylogenetic analysis was based on the taxon sampling used previously<sup>16</sup>. All 13 protein coding regions of the *Xenoturbella* mitochondrial genomes were extracted and translated into amino-acid sequences in Geneious using translation table 5 (invertebrate mitochondrial). Translated amino-acid sequences from the other taxa were downloaded directly from GenBank (Supplementary Table 4). The complete mitochondrial genomes of the four new *Xenoturbella* species were uploaded to the MITOS web server for annotation using translation table 5 (<http://mitos.bioinf.uni-leipzig.de>).

The translated amino-acid sequences of each protein-coding gene were aligned independently using MAFFT<sup>34</sup> and the FFT-NS-i × 1000 algorithm. Each of these 13 alignments was then run on the GBLOCKS server with the least stringent settings to remove poorly aligned regions ([http://molevol.cmima.csic.es/castresana/Gblocks\\_server.html](http://molevol.cmima.csic.es/castresana/Gblocks_server.html))<sup>35</sup>. We used Geneious to concatenate the trimmed sequences into a single alignment of 2,519 positions.

The amino-acid data were analysed using RAXML 8.1.22 (ref. 20) with the data partitioned by gene (13 partitions) and with the GTR +  $\Gamma$  model. Clade support was assessed using 100 bootstrap pseudo-replicates. A PhyloBayes analysis of the data was also run using CAT + GTR +  $\Gamma$ , as done previously with a similar data set<sup>16</sup>, although the files available from that study contained only 12 of the 13 protein coding genes (*COIII* was missing) and was also missing a further 361 bp from cytochrome b for *Amphipholis*.

For assessing the relationships among the five *Xenoturbella* species, the nucleotide data were analysed using RAXML with GTR +  $\Gamma$ . Clade support was assessed using 100 bootstrap pseudo-replicates. Rooting was based on the topology found in the analysis of the metazoan mitochondrial genomes (Extended Data Fig. 7).

**Phylogenomic analyses.** Taxa included in the transcriptome analysis are listed in Supplementary Table 5. Agalma<sup>36</sup> is an analysis pipeline (<https://bitbucket.org/caseydwinn/agalma>) for phylogenetics that was used to construct alignments of orthologous gene sequences from raw sequence data.

Adaptor sequences, low quality reads and ribosomal RNA sequences were removed from the *X. profunda* raw Illumina and NCBI SRA data. The transcriptomes were assembled using Trinity (version r20140413p1)<sup>37</sup> from the cleaned reads. The UCSC genome browser was used to download transcribed DNA sequences from the *mrna.fa.gz* file of the taxa with sequenced genomes. The assemblies and genome data were translated into amino-acid sequences by longest open reading frame. All of the translated transcriptomes were loaded into the Agalma database and then sent through the 'post-assemble' and 'phylogeny' pipelines. An all-by-all blast was used to find homologous sequences across the species based on sequence similarity. These blast results were aligned and run through RAXML 8.0 to make gene trees. These gene trees were used to identify orthologues; non-homologous sequences were removed. Two supermatrices were constructed using the homologous protein alignments. The complete matrix had 7,878 genes and 37% average gene occupancy. The matrix used in the phylogenetic analysis was trimmed down to 1,178 genes with 70% average gene occupancy (Supplementary Table 5).

The 70% matrix was used in all phylogenetic analysis. RAXML was used to calculate maximum likelihood trees. The -m PROTGAMMAAUTO option was called to automatically test for the best-fitting amino-acid substitution model. This selected the LG model. However, the more complex GTR model was not included in this test. For this reason, we ran rapid bootstrap analyses (100 replicates) and searches for the best scoring ML tree using both the LG and GTR amino-acid substitution models and the  $\Gamma$  model of rate heterogeneity (PROTLGGTR or PROTGAMMAGTR). We also ran the suboptimal model WAG +  $\Gamma$ , as used in previous studies involving *Xenoturbella*<sup>11</sup>.

The Swofford–Olson–Waddell–Hillis (SOWH) test<sup>38</sup> evaluates statistical support for incongruent phylogenetic topologies. We used SOWHAT (<https://github.com/josephryan/sowhat>)<sup>21</sup> to perform SOWH tests by comparing the following best constrained trees: (1) with Xenacoelomorpha as sister to Deuterostomia (((*Ciona*, *Gallus*, *Homo*, *Branchiostoma*, *Anneissia*, *Saccoglossus*, *Strongylocentrotus*), (*Hofstenia*, *X. bocki*, *X. profunda*)), *Peripatopsis*, *Capitella*, *Macrodasya*, *Baseodiscus*, *Priapulid*, *Lottia*, *Phoronis*, *Loxosoma*, *Gnathostomula*, *Adineta*, *Drosophila*, *Mnemiopsis*, *Trichoplax*, *Amphimedon*, *Nematostella*, *Schmidtea*); or (2) with Xenacoelomorpha as sister to Ambulacraria within Deuterostomia ((((*Saccoglossus*, *Strongylocentrotus*, *Anneissia*), (*Hofstenia*, *X. profunda*, *X. bocki*)), *Branchiostoma*, *Homo*, *Gallus*, *Ciona*), *Peripatopsis*, *Schmidtea*, *Capitella*, *Macrodasya*, *Baseodiscus*, *Priapulid*, *Lottia*, *Phoronis*, *Loxosoma*, *Gnathostomula*, *Adineta*, *Drosophila*, *Mnemiopsis*, *Trichoplax*, *Amphimedon*, *Nematostella*); against the maximum likelihood tree topology recovered using GTR +  $\Gamma$ . SOWHAT was run with RaxML for 130 replicates using WAG +  $\Gamma$ . WAG was used instead of LG or GTR because of computational capacity restrictions. Astral II<sup>22</sup> was used to estimate a species tree given a set of unrooted gene trees. A total of 393 gene alignments with an average occupancy of 80% were input into Astral II (<https://github.com/smirarab/ASTRAL/>). The Astral II species tree was generated from the 393 gene trees with 100 replicates of bootstrapping. PhyloBayes<sup>19</sup> MPI was run on CIPRES (<https://www.phylo.org/>) for 4,460 generations using the CAT + GTR +  $\Gamma$ 4 options shown previously to be optimal<sup>16</sup>. The first 2,200 trees were discarded as burn-in and the posterior consensus was computed on the remaining 2,260 trees. All PhyloBayes analyses were conducted using the CIPRES Science Gateway<sup>26</sup> and all other phylogenetic analysis were conducted on Amazon Web services C4 Instances (AWS EC2).



To assess if the phylogenetic signal differed between slow- and fast-evolving proteins, submatrices were constructed from the 70% average gene occupancy matrix. First, ribosomal proteins were screened from the original 70% matrix by blasting to the Uniprot ribosomal protein database (<http://www.uniprot.org/docs/ribosomp>). This filtered data set consisted of 1,127 genes (51 ribosomal proteins were removed). A gene tree was made for each gene and ranked by evolutionary rate<sup>23</sup> using a custom python script from (<https://github.com/NathanWhelan/Order-genes-by-evolutionary-rate>). The ranked genes were divided into the fastest-evolving 50% and the slowest-evolving 50%. The alignments were concatenated and then used to construct a maximum likelihood tree using the PROTGAMMAGTR amino-acid substitution model in RAxML. Support was assessed with 100 bootstrap pseudo-replicates.

All alignments, SOWHAT constraint trees and partition information are available on the Dryad repository (<http://datadryad.org/resource/doi:10.5061/dryad.79dq1>).

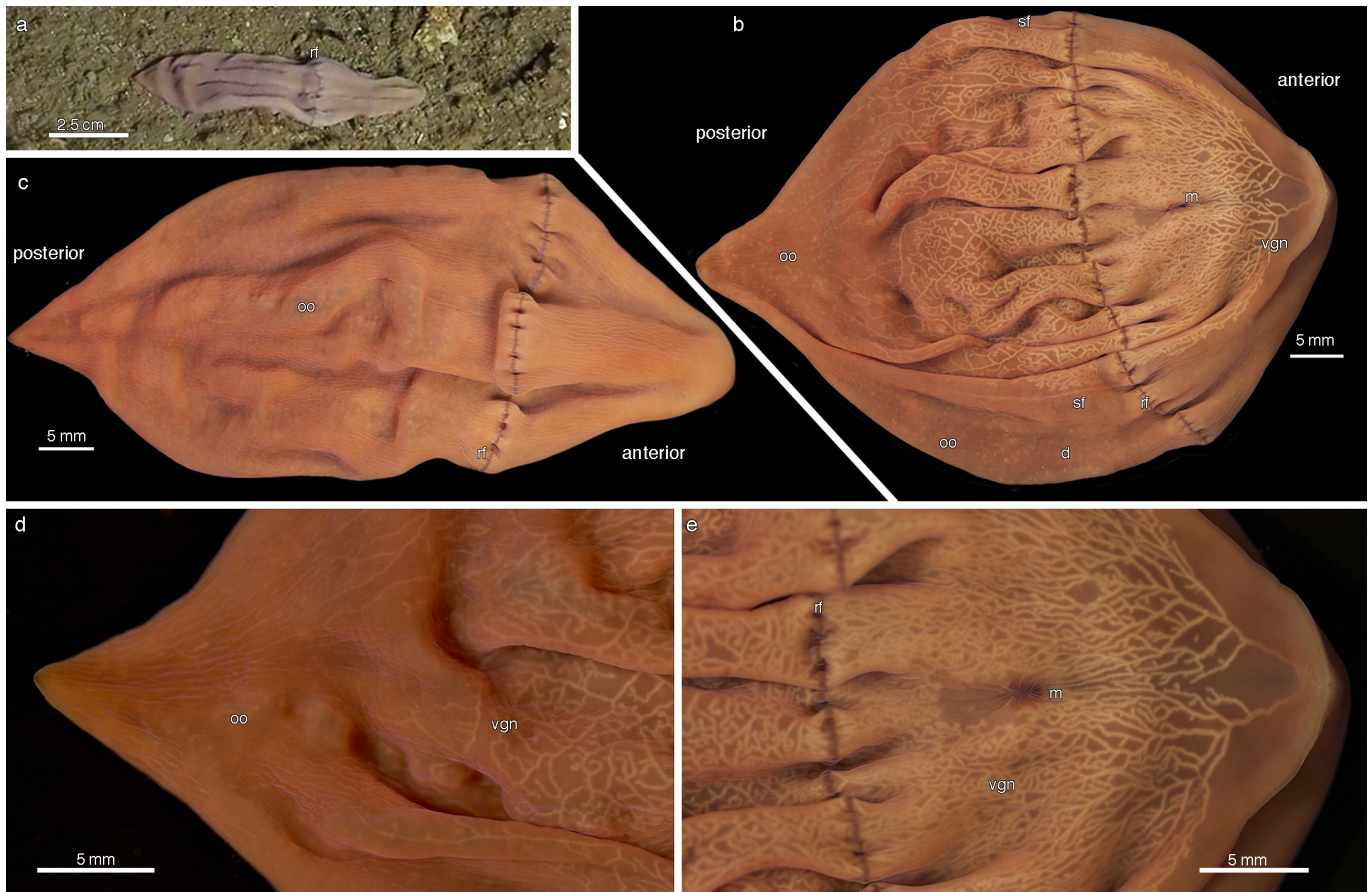
28. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
29. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291 (2007).
30. Boore, J. L. & Brown, W. M. Mitochondrial genomes of *Galathea*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Mol. Biol. Evol.* **17**, 87–106 (2000).
31. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
32. von Nickisch-Rosenegk, M., Brown, W. M. & Boore, J. L. Complete sequence of the mitochondrial genome of the tapeworm *Hymenolepis diminuta*: gene arrangements indicate that Platyhelminths are Eutrochozoans. *Mol. Biol. Evol.* **18**, 721–730 (2001).
33. Clement, M., Posada, D. & Crandall, K. A. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657–1659 (2000).
34. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).
35. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
36. Dunn, C. W., Howison, M. & Zapata, F. Agalma: an automated phylogenomics workflow. *BMC Bioinformatics* **14**, 330 (2013).
37. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol.* **29**, 644–652 (2011).
38. Goldman, N., Anderson, J. P. & Rodrigo, A. G. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* **49**, 652–670 (2000).



**Extended Data Figure 1 | *X. monstrosa* sp. nov.** **a**, Photograph taken by ROV of a vesicomyid clam field at ~3,000 m depth, Monterey Bay, California. Most clams are *E. extenta*. Two specimens of *Xenoturbella* are visible (arrows). Scale bar estimated from size of *E. extenta* (5 cm average width). **b**, Supplementary Video 1 frame grab showing same clam field as **a** with *A. diagonalis* (*Ad*) and *E. extenta* clams. Numerous *Xenoturbella* (arrows) were observed, including two specimens sampled for this study. Scale bar estimated from size of *E. extenta* (5 cm average width). **c**, Ventral view of the holotype SIO-BIC BI1037. Although highly contracted and incomplete (the posterior end was removed and frozen and the ventral area removed for histology), the specimen is still over 10 cm long. The mouth (*m*), ring furrow (*rf*) and side furrow (*sf*) are visible. **d**, Frame grab

of paratype SIO-BIC BI1039, a female *in situ*. **e**, Dorsal view of paratype SIO-BIC BI1039 (relaxed) showing ring furrow (*rf*), side furrow (*sf*), oocytes in body wall (*oo*) and part of the ventral surface (*v*). **f**, Ventral view of paratype SIO-BIC BI1039 showing the mouth (*m*), ring furrow (*rf*), oocytes in body wall (*oo*) and part of the epidermal ventral glandular network (*vgn*). **g**, Close-up of the ventral posterior of paratype SIO-BIC BI1039 showing the trailing off of the ventral glandular network (*vgn*) and oocytes clearly visible in the body wall. **h**, Close-up of the ventral anterior of paratype SIO-BIC BI1039 showing the mouth (*m*), ring furrow (*rf*) and the beginning of the ventral glandular network (*vgn*) near the anterior tip of the animal.

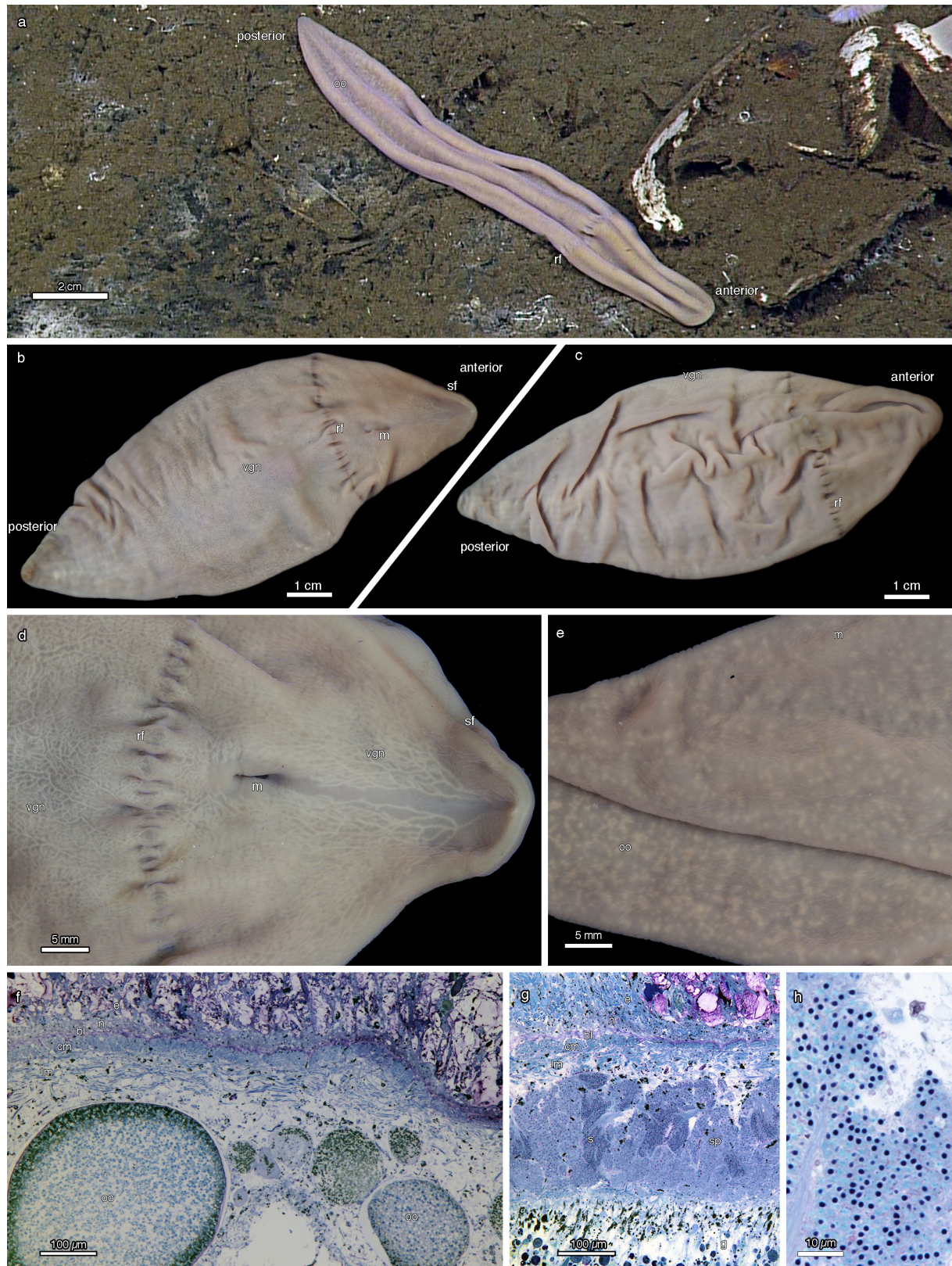




**Extended Data Figure 2 | *X. churro* sp. nov.** **a**, Frame grab of holotype SIO-BIC BI1040, a female, from the Guaymas Transform Fault, Gulf of California, Mexico, at ~1,700 m depth *in situ*. The ring furrow (rf) is visible towards the anterior end. **b**, Ventral view of the holotype SIO-BIC BI1040, (relaxed) showing mouth (m), ring furrow (rf), oocytes (oo), side furrow (sf) and epidermal ventral glandular network (vgn). Part of the dorsal side (d) is visible. **c**, Dorsal view of paratype SIO-BIC BI1039

showing ring furrow (rf) and oocytes (oo). **d**, Close-up of the ventral posterior of the holotype, showing the trailing off of the ventral glandular network (vgn), oocytes and the distinctively tapering posterior tip. **e**, Close-up of the anterior end of the holotype, showing the mouth (m), ring furrow (rf) and the beginning of the ventral glandular network (vgn) near the anterior tip of the animal.



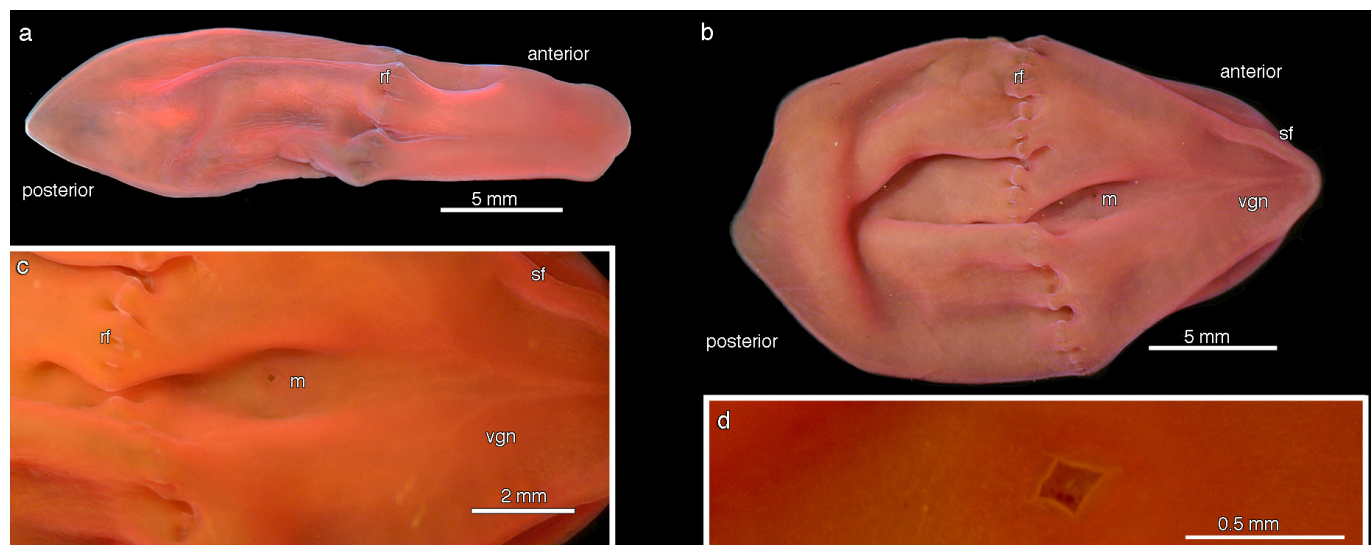


Extended Data Figure 3 | See next page for figure caption.



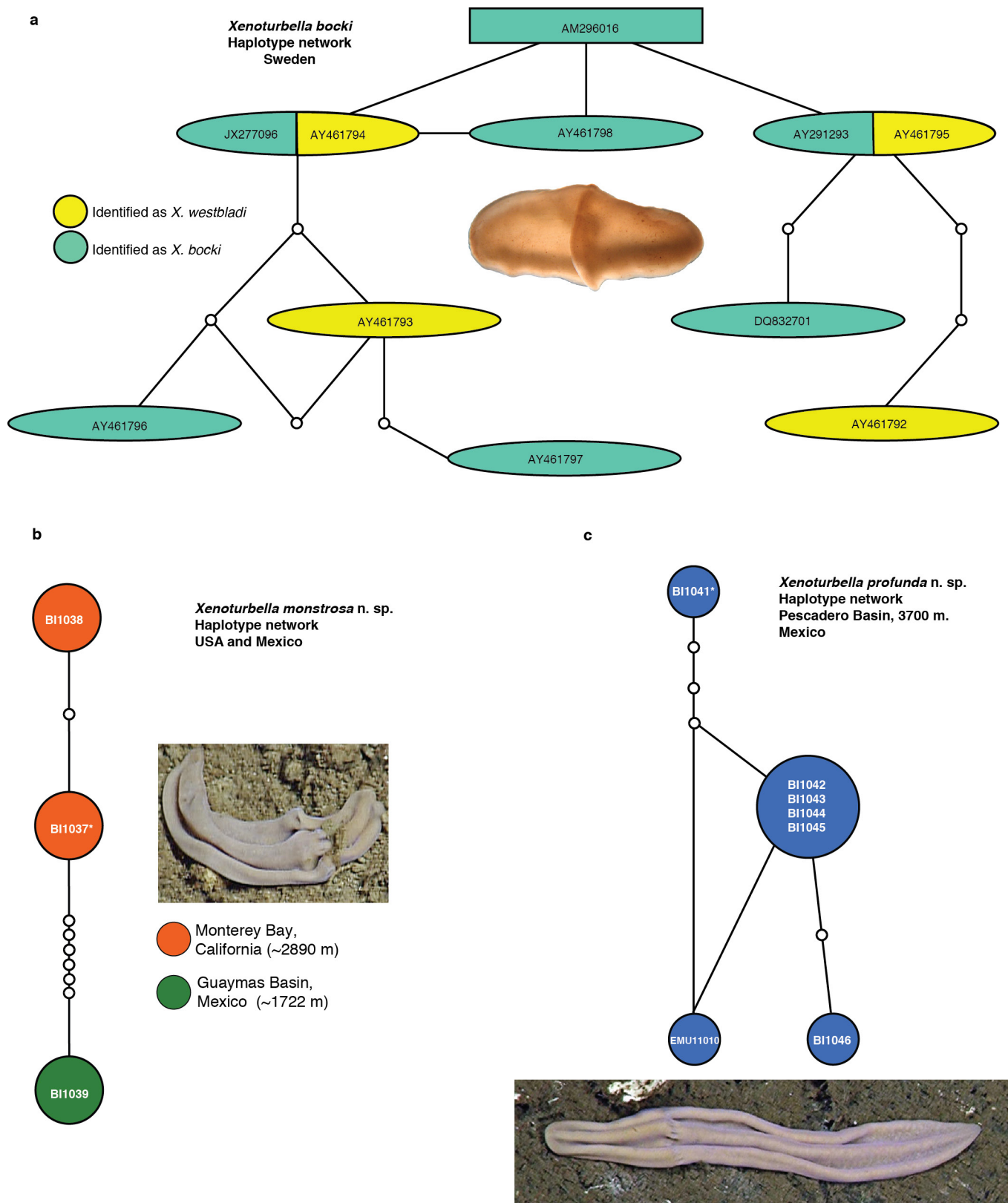
**Extended Data Figure 3 | *X. profunda* sp. nov.** **a**, Frame grab of paratype SIO-BIC BI1044, a female, from the Pescadero Basin, Gulf of California, Mexico, at ~3,700 m depth *in situ*. The ring furrow (rf) is visible towards the anterior end. **b**, Ventral view of the holotype SIO-BIC BI1041, a male (relaxed) showing mouth (m), ring furrow (rf), side furrow (sf) and epidermal ventral glandular network (vgn). **c**, Dorsal view of the holotype showing the ring furrow (rf). The ventral glandular network (vgn) is visible where part of the ventral side is exposed. **d**, Close-up of the ventral anterior end of female paratype SIO-BIC BI1044, showing mouth (m), ring furrow (rf), side furrow (sf) and beginning of ventral glandular network (vgn) near the anterior tip of the animal. **e**, Close-up of the dorsal posterior of female paratype SIO-BIC BI1044, showing oocytes of different

sizes distributed in the parenchyma. **f**, Parasagittal section (1  $\mu$ m, stained with toluidine blue) through dorsal body of female paratype SIO-BIC BI1044 showing thick epidermis (e), subepidermal nerve net (n), basal lamina (bl), circular muscle layer (cm), longitudinal muscle layer (lm) and parenchyma with oocytes of various sizes (oo). The gastrodermis is not shown here. **g**, Parasagittal section (1  $\mu$ m, stained with toluidine blue) through dorsal body wall of the male holotype SIO-BIC BI1041. Tissue layers as in the female, but a layer of developing (sp.) and mature sperm (s) instead of oocytes is present. The gastrodermis (g) lies beneath the parenchyma layer. **h**, Parasagittal section (1  $\mu$ m, stained with toluidine blue) showing parenchyma layer with late spermatids or mature sperm with spherical heads and free flagella.



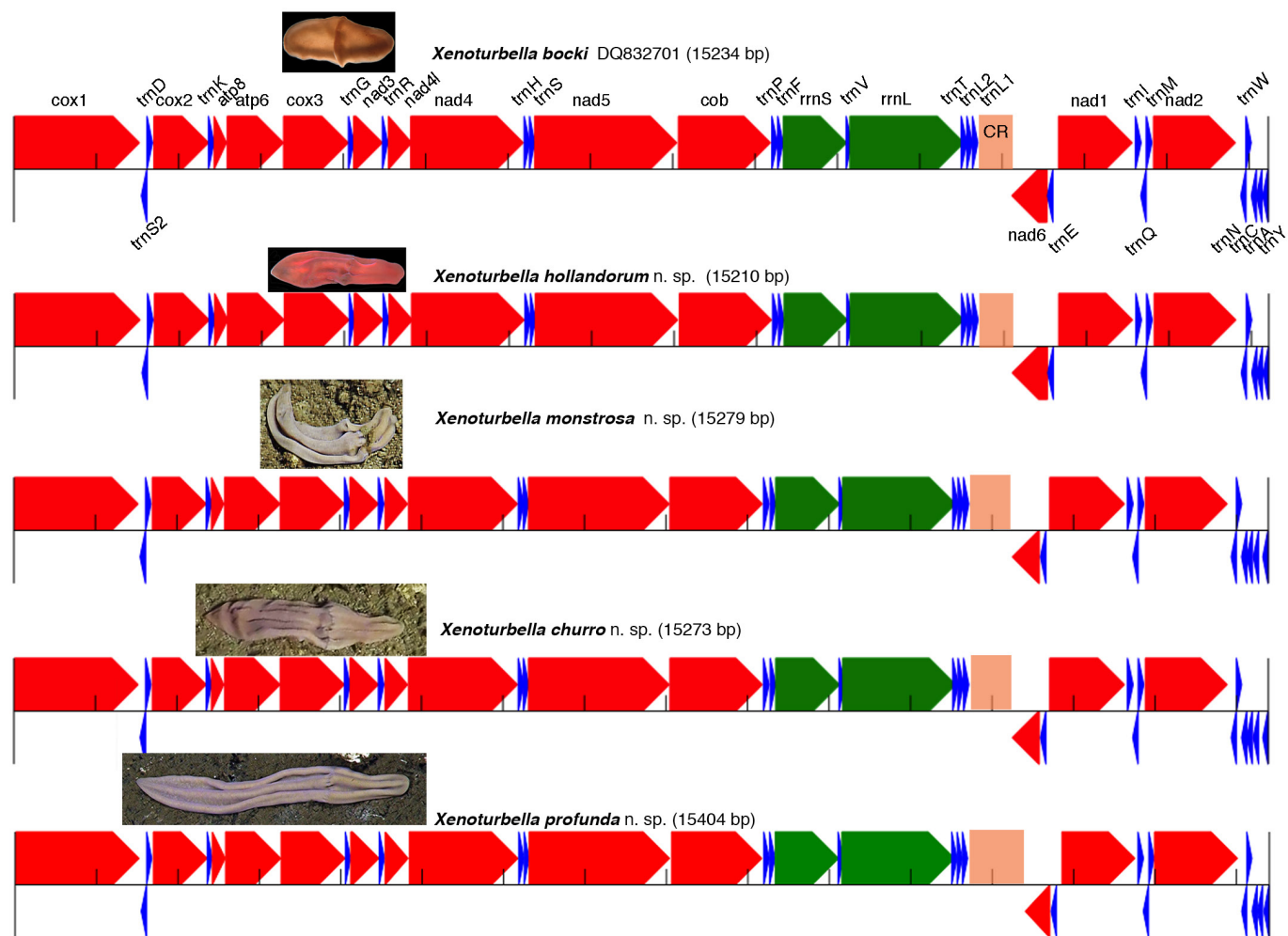
**Extended Data Figure 4** | *X. hollandorum* sp. nov. **a**, Dorsal view of the live unrelaxed holotype SIO-BIC BI1036, showing ring furrow (rf). **b**, Ventral view of the holotype (relaxed) showing mouth (m), ring furrow (rf) side furrow (sf) and epidermal ventral glandular network (vgn).

**c**, Close-up of the anterior end of the holotype, showing mouth (m), ring furrow (rf), side furrow (sf) and the ventral glandular network (vgn) near the anterior tip of the animal. **d**, Close-up of the diamond-shaped mouth.



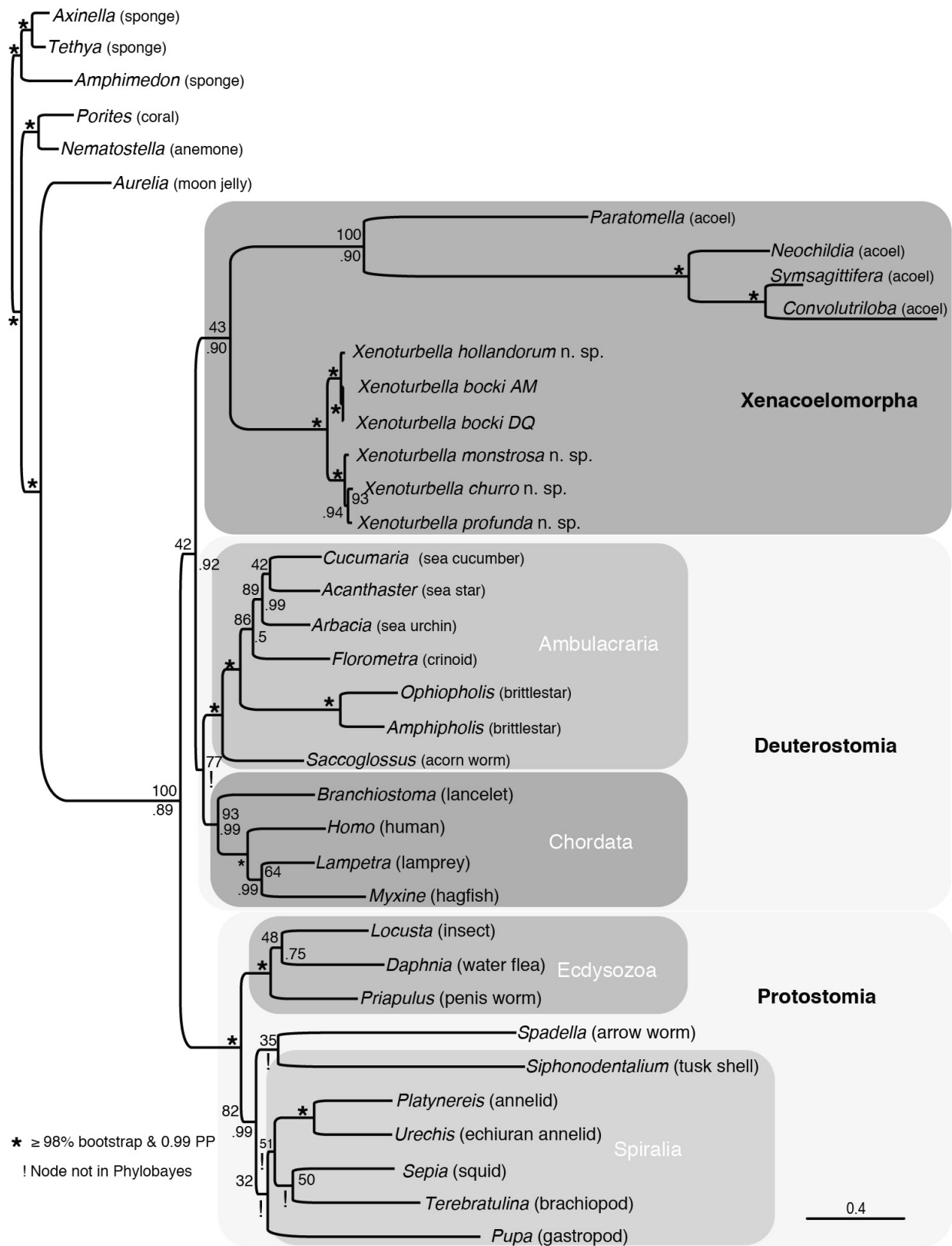
**Extended Data Figure 5 | Mitochondrial *COI* haplotype networks.**  
**a**, *X. bocki* and *X. westbladi* haplotypes (GenBank accession numbers indicated). Sequences from the two nominal species are completely interconnected. This network suggests that there is only one species in Swedish waters, which should be recognized under the older name, *X. bocki*. **b**, *X. monstrosa* from Monterey Bay (California) and Gulf of

California (Mexico) (SIO-BIC accession numbers indicated). **c**, *X. profunda* from the vicinity of a hydrothermal vent at 3,700 m in the Pescadero Basin, Mexico (SIO-BIC and UNAM specimens indicated). Holotypes are designated with asterisks in the figure. Networks generated with TCS<sup>33</sup>.



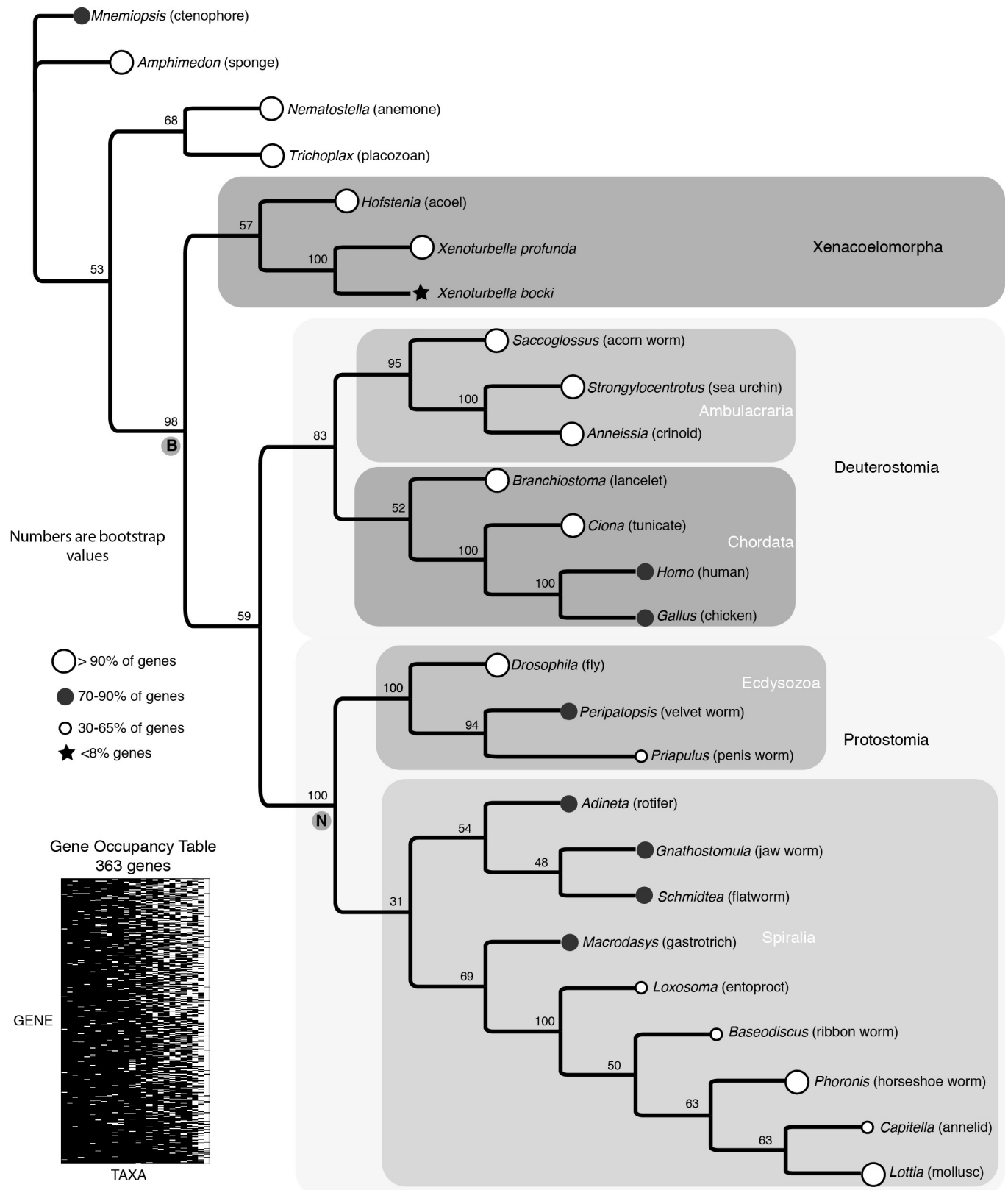
**Extended Data Figure 6 | Mitochondrial genomics of *Xenoturbella*.** a, Gene order for the four new species compared with that of *X. bocki*<sup>9,18</sup>. The gene order was consistent across all species with only minor variation in amino acids and length of the control region (also see Supplementary Table 1).



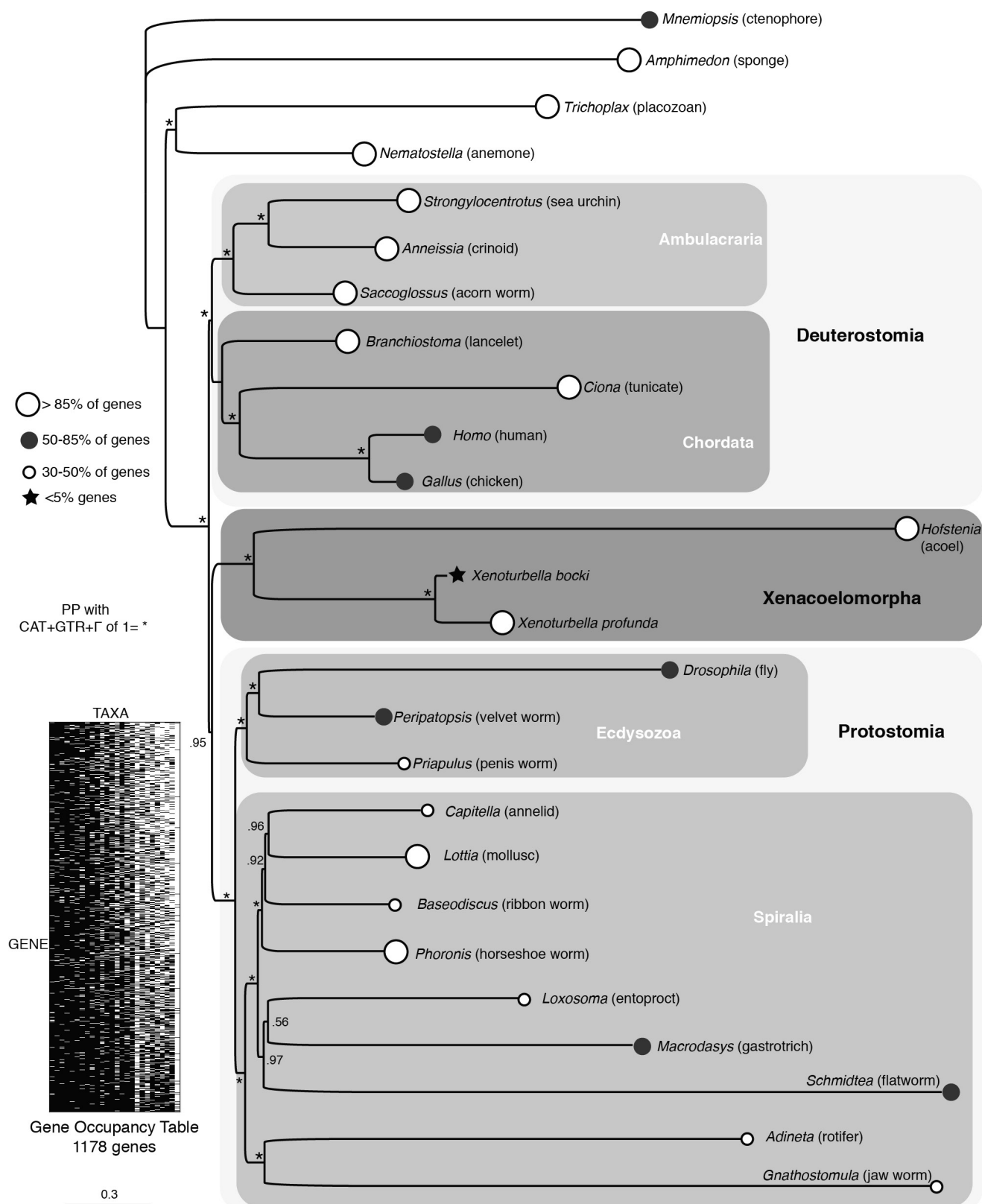


**Extended Data Figure 7 | Maximum likelihood phylogeny based on all 13 mitochondrial proteins.** Five *Xenoturbella* species formed a weakly supported sister clade to Acoela = Xenacoelomorpha. Xenacoelomorpha was sister to deuterostomes with weak support. Data were partitioned and analysed using RAXML and GTR +  $\Gamma$ . Analysis using PhyloBayes under CAT + GTR +  $\Gamma$  placed Xenacoelomorpha inside deuterostomes as a weakly supported sister group to Chordata (not shown), a result not

recovered previously<sup>16</sup> with the same program, model and terminals (except for *X. profunda*). Scale bar, substitutions per amino-acid position. The lower support for Xenacoelomorpha in its placement with deuterostomes, or its placement with Chordata, compared with previously shown, may be due to additional *Xenoturbella* data and/or the fact that all 13 mitochondrial proteins were used here.

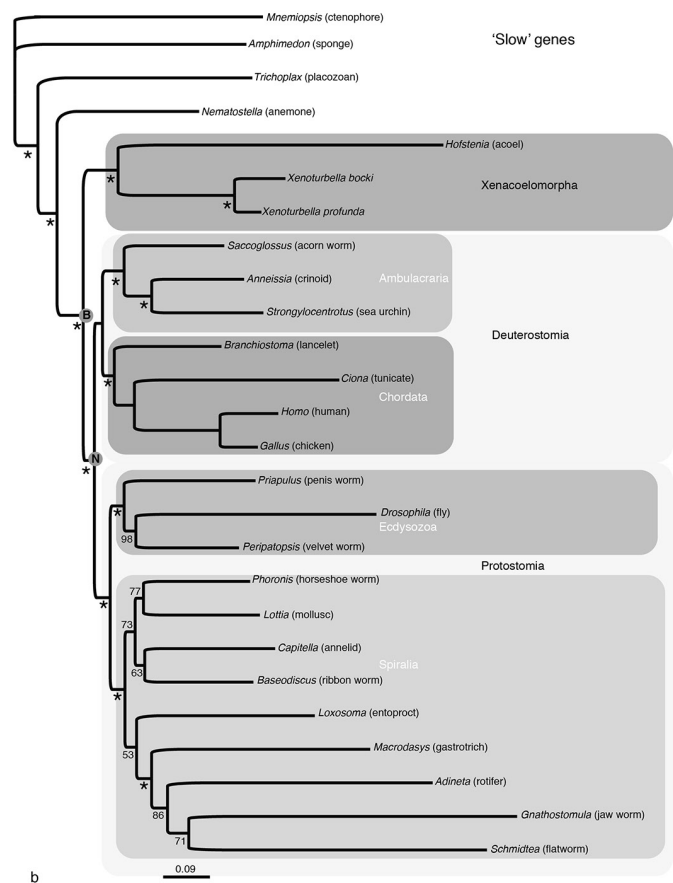
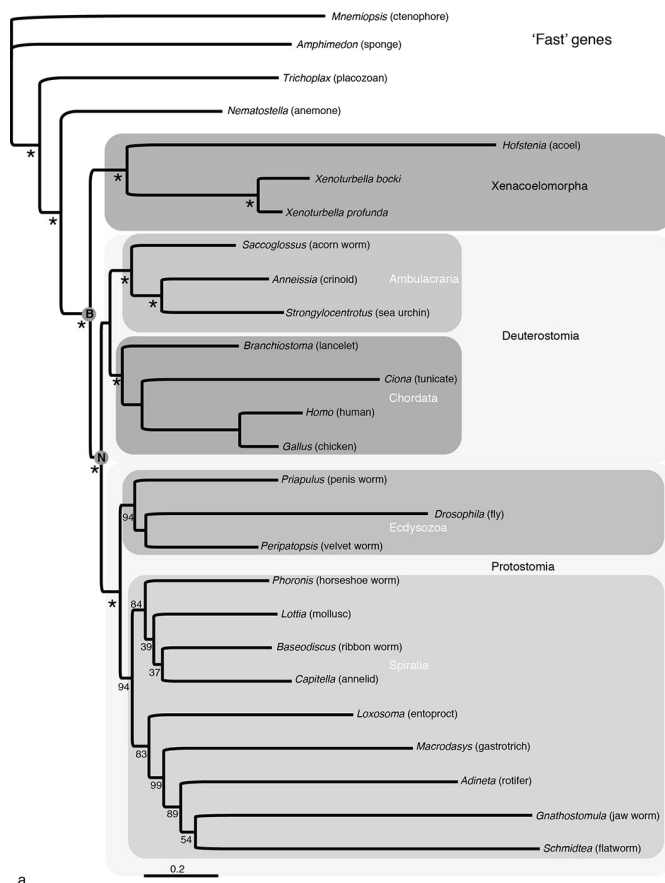


**Extended Data Figure 8 | Species tree based on 393 gene trees using Astral II<sup>22</sup>.** Nodal support assessed via 100 pseudo-replicates of bootstrapping for each gene, combining both information from each gene tree and its respective bootstrap replicates. B, Bilateria; N, Nephrozoa.



**Extended Data Figure 9 | PhyloBayes analysis of 1,178 genes with 70% average gene occupancy.** CAT + GTR +  $\Gamma$  was used and the first 2,000 out of 4,460 generations for two chains were discarded as burn-in and the remainder used to generate a consensus tree. Scale bar, substitutions per

amino-acid position. The largest discrepancy observed between chains was 0.164609 and the mean was 0.000989. Xenacoelomorpha grouped with protostomes. B, Bilateria.



**Extended Data Figure 10 | Maximum likelihood analyses of 'fast'- and 'slow'-evolving data sets.** The 70% occupancy data set was reduced to 1,127 genes (51 ribosomal proteins removed) and then divided into the fastest-evolving 50% and the slowest-evolving 50%. These were analysed using the PROTAMMAGTR amino-acid substitution model in RAXML.

**a**, Phylogeny based on the 'fast-evolving' data set with 216,066 amino acids. **b**, Phylogeny based on the 'slow-evolving' data set with 168,571 amino acids. Nodal support assessed via 100 pseudo-replicates of bootstrapping. Asterisks indicate 100% bootstrap support. B, Bilateria; N, Nephrozoa.



# Autism-like behaviours and germline transmission in transgenic monkeys overexpressing MeCP2

Zhen Liu<sup>1\*</sup>, Xiao Li<sup>1\*</sup>, Jun-Tao Zhang<sup>1</sup>, Yi-Jun Cai<sup>1</sup>, Tian-Lin Cheng<sup>1</sup>, Cheng Cheng<sup>1</sup>, Yan Wang<sup>1</sup>, Chen-Chen Zhang<sup>1</sup>, Yan-Hong Nie<sup>1</sup>, Zhi-Fang Chen<sup>1</sup>, Wen-Jie Bian<sup>1</sup>, Ling Zhang<sup>2</sup>, Jianqiu Xiao<sup>2</sup>, Bin Lu<sup>1</sup>, Yue-Fang Zhang<sup>1</sup>, Xiao-Di Zhang<sup>1</sup>, Xiao Sang<sup>1</sup>, Jia-Jia Wu<sup>1</sup>, Xiu Xu<sup>3</sup>, Zhi-Qi Xiong<sup>1</sup>, Feng Zhang<sup>2</sup>, Xiang Yu<sup>1</sup>, Neng Gong<sup>1</sup>, Wen-Hao Zhou<sup>4</sup>, Qiang Sun<sup>1</sup> & Zilong Qiu<sup>1</sup>

Methyl-CpG binding protein 2 (MeCP2) has crucial roles in transcriptional regulation and microRNA processing<sup>1–4</sup>. Mutations in the *MECP2* gene are found in 90% of patients with Rett syndrome, a severe developmental disorder with autistic phenotypes<sup>5</sup>. Duplications of *MECP2*-containing genomic segments cause the *MECP2* duplication syndrome, which shares core symptoms with autism spectrum disorders<sup>6</sup>. Although *Mecp2*-null mice recapitulate most developmental and behavioural defects seen in patients with Rett syndrome, it has been difficult to identify autism-like behaviours in the mouse model of MeCP2 overexpression<sup>7,8</sup>. Here we report that lentivirus-based transgenic cynomolgus monkeys (*Macaca fascicularis*) expressing human MeCP2 in the brain exhibit autism-like behaviours and show germline transmission of the transgene. Expression of the *MECP2* transgene was confirmed by western blotting and immunostaining of brain tissues of transgenic monkeys. Genomic integration sites of the transgenes were characterized by a deep-sequencing-based method. As compared to wild-type monkeys, *MECP2* transgenic monkeys exhibited a higher frequency of repetitive circular locomotion and increased stress responses, as measured by the threat-related anxiety and defensive test<sup>9</sup>. The transgenic monkeys showed less interaction with wild-type monkeys within the same group, and also a reduced interaction time when paired with other transgenic monkeys in social interaction tests. The cognitive functions of the transgenic monkeys were largely normal in the Wisconsin general test apparatus, although some showed signs of stereotypic cognitive behaviours. Notably, we succeeded in generating five F<sub>1</sub> offspring of *MECP2* transgenic monkeys by intracytoplasmic sperm injection with sperm from one F<sub>0</sub> transgenic monkey, showing germline transmission and Mendelian segregation of several *MECP2* transgenes in the F<sub>1</sub> progeny. Moreover, F<sub>1</sub> transgenic monkeys also showed reduced social interactions when tested in pairs, as compared to wild-type monkeys of similar age. Together, these results indicate the feasibility and reliability of using genetically engineered non-human primates to study brain disorders.

We first co-injected lentivirus expressing synapsin-promoter-driven<sup>10</sup> haemagglutinin (HA)-tagged human MeCP2 and green fluorescence protein (GFP) and lentivirus expressing mCherry into the perivitelline space of 94 mature oocytes of cynomolgus monkeys (Fig. 1a). We found that 61 out of 88 (69%) of the surviving oocytes became zygotes after intracytoplasmic sperm injection (ICSI), and 53 embryos were then transferred into 18 surrogate monkeys. Nine surrogates (9 out of 18, 50%) became pregnant and produced eight live births (3 male, 5 female; Fig. 1b) and four stillbirths, all carrying human *MECP2*, GFP and mCherry transgenes, as determined by PCR (Fig. 1c). The AccuCopy assay showed that the copy numbers of *MECP2*

transgenes in 8 live (T04–T11) and 2 aborted (T01 and T02) transgenic (TG) monkeys varied from 1.0 to 7.3 (Extended Data Table 1a). In the second experiment, we injected 264 mature oocytes with lentivirus carrying the hSynapsin-HA-hMECP2-2a-GFP cassette, and transferred 105 embryos after ICSI into 36 surrogates. Owing to unfavourable seasonal conditions, only 7 pregnant surrogates gave birth to 9 monkeys (T13–T21), and only 2 survived (Supplementary Table 1).

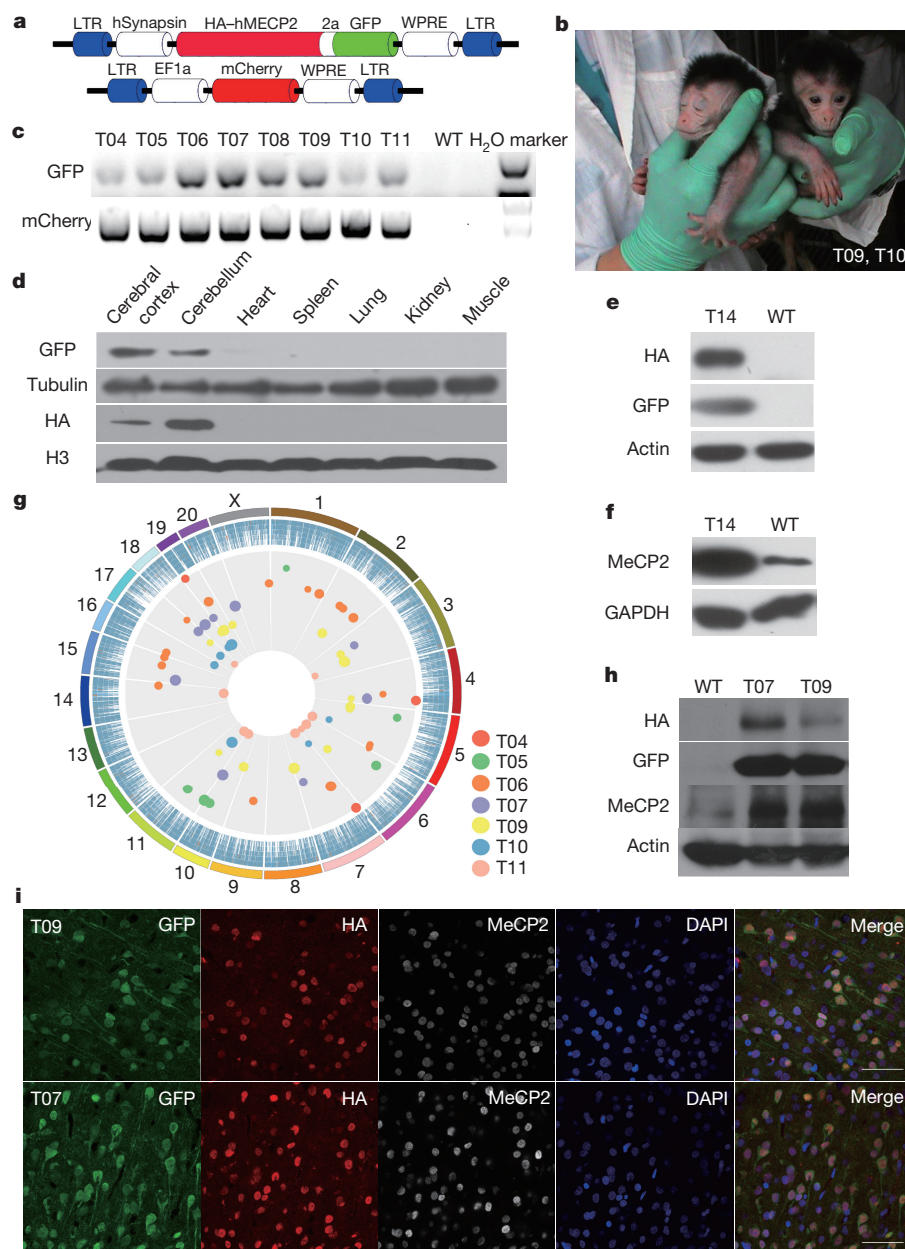
Western blotting of tissues of stillbirth TG monkey T14 showed expression of GFP and HA-MeCP2 proteins in the cortex and cerebellum, but not in non-neural tissues, confirming specific transgene expression under the synapsin promoter (Fig. 1d). Levels of MeCP2 protein were also significantly higher than that found in an aborted wild-type (WT) monkey of a similar age (Fig. 1e, f). Transgenic integration was confirmed by Southern blotting using a probe targeting the HA-hMECP2-2a-GFP transgene (Extended Data Fig. 1a). We next analysed genomic integration sites of lentiviral cassettes containing HA-hMECP2-2a-GFP and mCherry transgenes by a deep-sequencing-based method (Extended Data Fig. 1b). All transgenes were located in genomic loci distant from known coding exons, and thus unlikely to interfere with endogenous genes (Fig. 1g and Supplementary Table 2), and insertion numbers were largely consistent with the copy numbers identified by AccuCopy (Extended Data Fig. 1c). Therefore, the human *MECP2* transgene was successfully incorporated into the monkey genome and specifically expressed in the monkey's brain.

Compared to WT monkeys of similar ages, the body weight and abdominal circumference of the TG group (T04–T11) was slightly lower before 20 months of age, whereas no difference was found for head-trunk length, heart and respiratory rates or body temperature (Extended Data Fig. 2a–g). We did not observe in TG monkeys any seizure phenotype, which was associated with *MECP2* duplication syndrome patients<sup>6</sup>, perhaps owing to the young age of the monkeys. Interestingly, mass spectrometry of blood metabolites at ~18 and ~36 months suggested metabolic abnormalities in the TG group, with significantly higher levels of some short- and long-chain fatty acids (Extended Data Fig. 3a, b), reminiscent of some human autistic patients<sup>11</sup>.

Despite their generally normal early development, one TG monkey (T05) showed severe weight loss and head circumference reduction after 15 months (Extended Data Fig. 4a–c), and was unable to complete behavioural tests. Monkeys T09 and T07 became severely sick at 43 and 46 months, respectively, after behavioural tests. The sickness of these TG monkeys echoed the severe phenotypes of human patients with the *MECP2* duplication syndrome<sup>6</sup>. The euthanasia procedure was performed, and their brain tissues were collected for further analysis with western blotting, immunostaining and RNA-sequencing (RNA-seq). We found that HA-MeCP2 and GFP were expressed in the brain

<sup>1</sup>Institute of Neuroscience, CAS Key Laboratory of Primate Neurobiology, State Key Laboratory of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, China. <sup>2</sup>State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200438, China. <sup>3</sup>Department of Child Healthcare, Children's Hospital of Fudan University, Shanghai 201102, China. <sup>4</sup>Department of Neonatology, Children's Hospital of Fudan University, Shanghai 201102, China.

\*These authors contributed equally to this work.



**Figure 1 | Construction of *MECP2* transgenic monkey and brain-specific expression of transgenes.** **a**, Top, lentiviral HA-hMECP2-2a-GFP cassette. Bottom, the EF1a-mCherry cassette. **b**, Image of newborn *MECP2* transgenic (TG) monkeys (T09 and T10). Photo credit: Y.W. **c**, PCR analysis showing the presence of transgenes (GFP, top; mCherry, bottom) in 8 live TG monkeys' genomes. **d**, Brain-specific transgene expression, shown by western blots of different tissues of T14. Top two panels: cytosolic fractions; bottom two panels: nuclear fractions, stained with antibodies indicated. Note the exposure times for transgene and the loading control were different (see Supplementary Fig. 1 for further

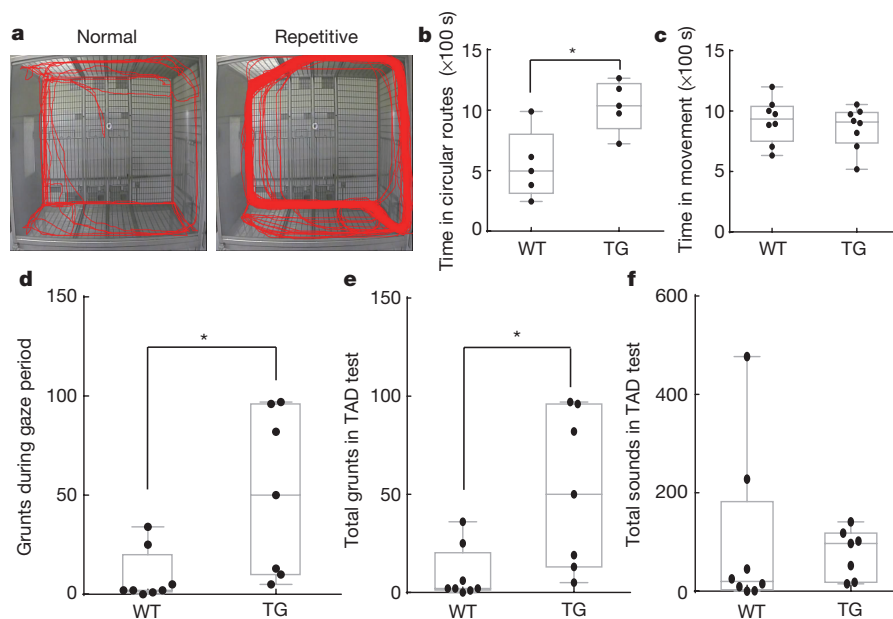
comparison). H3, histone 3. **e**, **f**, **h**, Western blots showing expression of HA-MeCP2 and GFP in brain tissues of TG (T07, T09 and T14) and wild-type (WT) monkeys. For gel source data, see Supplementary Figs 1–3. **g**, Genome-wide distribution of transgenes in F<sub>0</sub> TG monkeys. Insertion sites (dots) distribute on various chromosomes (outermost circle). Sizes of dots are proportional to reads identified by deep-sequencing, colour-coded and aligned circularly for different monkeys (sample for T08 absent owing to preparation failure). **i**, Immunostaining of cortical sections of brains of T07 and T09 for GFP, HA, MeCP2 and DAPI. Scale bars, 50  $\mu$ m.

lysates of T07 and T09 (Fig. 1h). Immunostaining of cortical slices of T07 and T09 showed that the MeCP2 and HA signals were co-localized (Fig. 1i), indicating expression of the *MECP2* transgene in the TG monkeys' brain.

Further transcriptome-wide analysis of the brain tissues was performed on four deceased TG (T14, T05, T07 and T09) and four WT monkeys using RNA-seq, based on the whole-genome sequencing data for cynomolgus monkeys<sup>12–14</sup>. We found 105 upregulated and 209 downregulated genes in TG monkeys (Extended Data Fig. 4d, e), with  $\geq 2$ -fold change as compared to WT monkeys. Among them, 13 upregulated and 3 downregulated genes were

also reported to exhibit similar changes in the *MECP2* transgenic mice<sup>15</sup>.

Motor functions and responses to stress<sup>16–18</sup> were examined for 8 TG (T04–T11) and 8 WT monkeys (aged 12–18 months). First, we video-recorded the locomotion of each monkey alone for 20 min per day for 5 days, and found that four TG (T04, T05, T06 and T09) and two WT monkeys exhibited repetitive circular locomotion (in the same direction, at least three times; Supplementary Videos 1 and 2). The total time spent in circular locomotion during the observation period (average over 5 days) for all eight transgenic monkeys was significantly higher than that of eight WT monkeys (Fig. 2a, b). This difference



**Figure 2 | Alterations in locomotion activity and increased anxiety in *MECP2* TG monkeys.** **a**, Examples of movement trajectories (red traces) viewed from cage-top, showing normal activity (left) and repetitive circular routing (right). **b**, Boxplots of time spent in repetitive routing of TG and WT monkeys ( $n = 8$  each), monitored for 20 min daily for 5 days. Each dot depicts data from 1 day ( $*P = 0.014$ , Student's  $t$ -test). **c**, Average total time spent in movements in 20-min period ( $n = 8$  each). **d**, Results from threat-related anxiety and defence (TAD) test. Boxplots of numbers

of grunts during the gaze period ( $n = 8$ , WT;  $n = 7$ , TG;  $*P = 0.009$ , Mann-Whitney  $U$  test) at 18 months of age. **e**, **f**, Total number of grunts (**e**) and all sounds (**f**) during the entire TAD tests (same monkey sets as in **d**;  $*P = 0.014$ , Mann-Whitney  $U$  test). Ends of whiskers represent the minimum and maximum of data points. The line within box represents the median (odd numbers of data points) or second quartile (even number of data points). The bottom and top edge of box represents the first and the third quartile, respectively.

was not due to hyperactivity of the TG monkeys, because the total time spent in locomotion was similar between the two groups during observation (Fig. 2c).

Anxiety-associated behaviours were found in autism patients and mouse models of *MECP2* overexpression<sup>6,7</sup>. We used the threat-related anxiety and defensive (TAD) behavioural test<sup>9</sup> to assay the vocalization responses of the monkeys to human gaze (Extended Data Fig. 5a). Typical sounds include grunt, coo and scream (Supplementary Audio 1–3), each was identified by its distinct signature in the sound spectrograph (Extended Data Fig. 6a–c). Notably, we found that at ~18 months of age, the total number of grunts made by the TG group during the gaze period and the entire TAD test was significantly higher than that of the WT group (Fig. 2d, e), with similar total numbers of sounds (grunt, coo and scream) per session produced by both groups (Fig. 2f and Supplementary Table 3). This increase in anxiety-related grunts was also found at 36 months of age (Extended Data Fig. 5b–d). Thus, *MECP2* TG monkeys showed increased levels of anxiety.

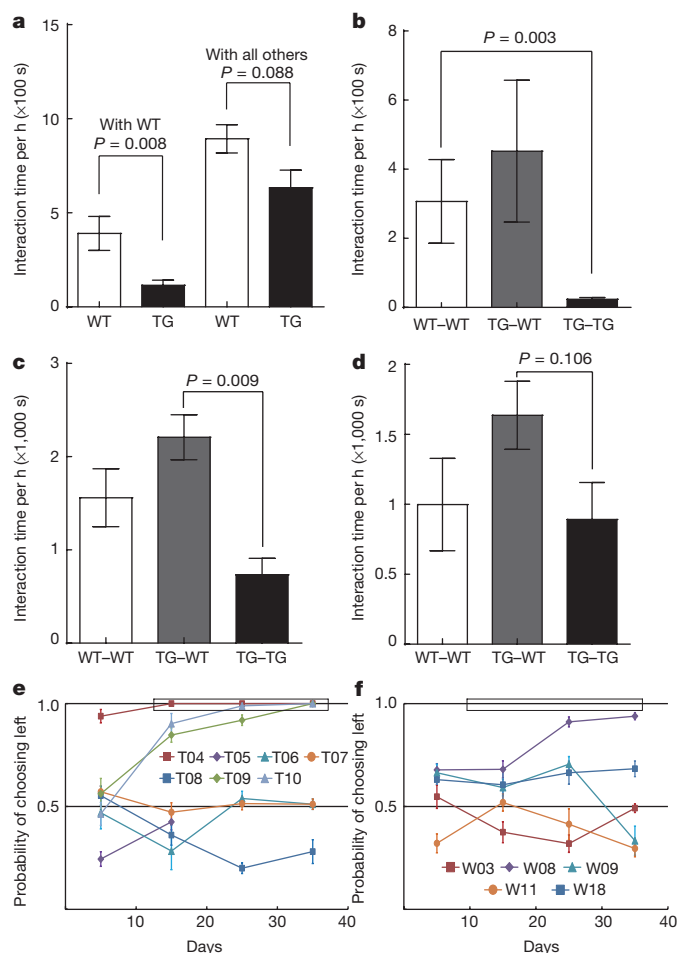
Impairment of social interaction is a hallmark of autism and *MECP2*-associated disorders. We examined the time monkeys sat together with apparent interactions, a prominent social behaviour in monkey colonies<sup>19–21</sup>. First, three groups of monkeys (at ~18 months of age, two TG and three WT in each group) were reared together for over 6 months (Supplementary Table 4a). We found that the average time a TG monkey sat together with another WT monkey within the group was significantly lower than that of the WT monkey (60-min daily observation for 5 days, Fig. 3a and Supplementary Video 3; no data for TG–TG interaction owing to the limited number of TG monkeys). The total time all TG monkeys sat with any other monkey (either TG or WT) was also slightly lower than for WT monkeys (Fig. 3a). Next (at ~24 months of age), we paired two female monkeys from different groups in a single cage (60 min daily for 5 days), and found that the interaction time of TG–TG pairs was significantly lower than that of WT–WT pairs (Fig. 3b, Supplementary Video 4 and Supplementary Table 4b). Pairing of unfamiliar male monkeys was not performed owing to their aggressive behaviours near adolescence. Finally

(at 36 months of age), we paired familiar female and male monkeys from the same group, in which male pairs showed no aggressive interaction, and found that the male TG–TG pairs interacted less than that of the TG–WT pairs, whereas the difference between female TG–TG pairs and TG–WT pairs was not significant (Fig. 3c, d, Extended Data Fig. 7a–f and Supplementary Table 4c). All of these social interaction tests were performed with observers blinded to monkey genotypes. This apparent difference between male and female TG monkeys is reminiscent of the finding that *MECP2* duplication syndrome show more severe autism-related symptoms in male patients<sup>6</sup>.

Cognitive function tests were performed using the Wisconsin general test apparatus (WGTA)<sup>22–26</sup>. During adaptation, discrimination and reversal steps of black/white tests, both the WT and TG groups passed each step with a similar average time course, but the TG monkeys exhibited much larger variability and one (T11) failed to pass the black/white reversal step (and was thus dropped from subsequent tests) (Extended Data Fig. 8a, b and Supplementary Table 5). In the Hamilton searching tests (adaptation, searching, set-breaking and forced set-breaking), the TG group showed a slightly slower learning in the forced set-breaking step (Extended Data Fig. 8c, d). Finally, the two groups showed no significant difference in the average performance in the reward-shape association learning test (Extended Data Fig. 9a–c and Supplementary Video 5). However, three out of seven TG monkeys showed marked left-side preference, regardless of the left or right location of the reward, a behaviour that was not observed in the WT monkeys (Fig. 3e, f and Supplementary Video 6). Thus, *MECP2* transgene expression resulted in some abnormalities in cognitive behaviours.

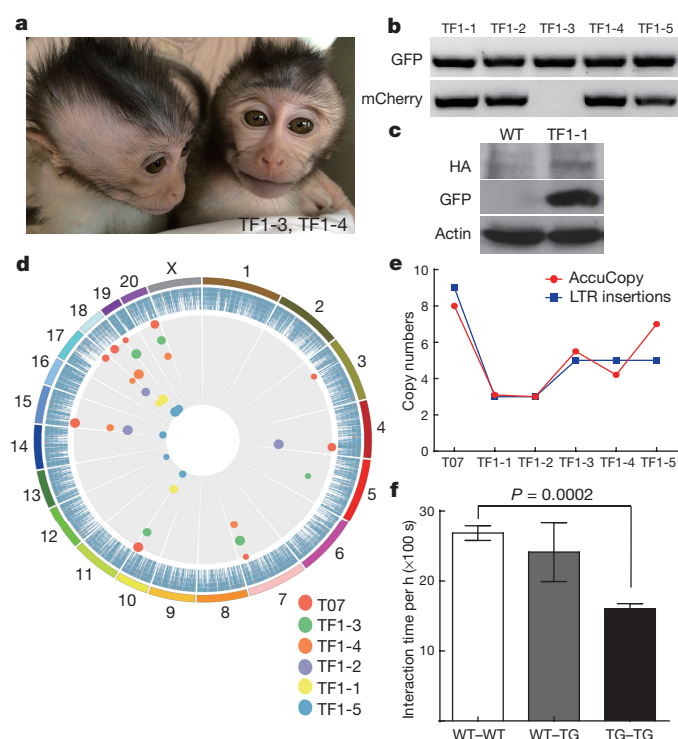
We further examined the germline transmission of *MECP2* TG monkeys, in view of previous lentiviral-based transgenic monkey experiments<sup>27,28</sup>. To facilitate the reproduction of TG monkeys, we used a recently developed testicular tissue xenografting method (see Methods). In brief, one testicle was obtained from T07 at 27 months of age, and pieces of the testicle tissue were xenografted subcutaneously in nude mice. Mature motile sperm were obtained from the xenografts after 10 months and used for ICSI on 176 monkey oocytes.





**Figure 3 | Impaired social interaction and cognitive functions in *MECP2* TG monkeys.** **a**, Interaction at 18 months of age among monkeys within the same group ('familiar', reared together for 6 months), defined by the average time a WT or a TG monkey sat together (for  $>3$  s) with a WT monkey from the same group (left), or with any other monkey (right) during a 60-min observation period ( $n = 6$ , TG;  $n = 9$ , WT). **b**, Interaction time between a pair of female monkeys from a different group ('unfamiliar', reared separately for 12 months) at 24 months of age, for WT-WT, WT-TG and TG-TG pairs ( $n = 4$ , TG;  $n = 12$ , WT). **c**, **d**, Interaction time between familiar monkey pairs at 36 months of age (c, male pairs,  $n = 3$ , TG;  $n = 8$ , WT; d, female pairs,  $n = 4$ , TG;  $n = 11$ , WT) from the same group reared together for 6 months. All  $P$  values are from Mann-Whitney  $U$  test. **e**, **f**, Data from WGTA test ( $n = 7$ , TG;  $n = 5$ , WT), showing distinct left bias in the probability of left or right choice in three TG monkeys. Boxed areas denote  $>95\%$  probability of choosing left side. Error bars denote s.e.m.

Implantation of 95 fertilized zygotes into 22 surrogates resulted in five F<sub>1</sub> offspring (TF1-1 to TF1-5; 4 live births; Fig. 4a and Extended Data Table 1b). The PCR analysis showed that all five F<sub>1</sub> monkeys carried the human *MECP2* transgene (Fig. 4b), and western blotting showed the expression of HA-MeCP2 and GFP in brain lysates of TF1-1 (deceased 3 days after birth) (Fig. 4c). Thus, the transgenes were expressed in the F<sub>1</sub> offspring. Deep-sequencing further showed the genomic integration sites of the HA-hMECP2-2a-GFP and mCherry transgenes in F<sub>1</sub> TG monkeys (Fig. 4d and Supplementary Table 6). As expected, transgenes in F<sub>1</sub> TG monkeys (TF1-1 to TF1-5) were mostly distributed in chromosomes that were a subset of transgene-containing chromosomes of the paternal monkey T07 (Fig. 4d), showing Mendelian segregation of transgenes among the F<sub>1</sub> progeny. Moreover, long-terminal repeat (LTR) insertion sites were consistent with AccuCopy-identified copy numbers, confirming the germline transmission of transgenes (Fig. 4e).



**Figure 4 | Generation of F<sub>1</sub> progeny of *MECP2* TG monkey.** **a**, Image of two newborn F<sub>1</sub> TG monkeys (TF1-3 and TF1-4). Photo credit: Y.-H.N. **b**, PCR analysis showing the presence of transgenes (GFP and mCherry) in five F<sub>1</sub> monkeys' genomes. **c**, Western blots showing expression of the HA-MeCP2 and GFP in F<sub>1</sub> monkeys' brain. For gel source data, see Supplementary Fig. 4. **d**, Genomic-wide distribution of transgenes in T07 and five F<sub>1</sub> offspring. Insertion sites (dots) distribute on various chromosomes (outermost circle). Sizes of dots are proportional to reads identified by deep-sequencing, colour-coded and aligned circularly for different monkeys. **e**, Copy numbers obtained with AccuCopy (red, *MECP2* and mCherry) are consistent with LTR insertion numbers identified by deep-sequencing (blue). **f**, Social interaction time between *MECP2* TG F<sub>1</sub> and WT monkeys from a different group ('unfamiliar pairing') at 11 months of age, for WT-WT, WT-TG and TG-TG pairs ( $P$  value, Student's  $t$ -test) ( $n = 4$ , TG;  $n = 6$ , WT). Error bars denote s.e.m.

Finally, we examine whether F<sub>1</sub> TG monkeys may also show defects in social interaction. We set up two groups of monkeys at 11 months of age (two TG and three WT in each group; Supplementary Table 7), then paired two monkeys from two different groups in a single cage (60 min daily for 5 days), observed in a blinded manner. We found that the TG-TG pairs showed a significantly lower total interaction time than the WT-WT pairs (Fig. 4f), although in general young monkeys exhibited more frequent interaction than older monkeys. Thus, defective social behaviours were inherited in the F<sub>1</sub> generation of *MECP2* TG monkeys.

In summary, we have generated transgenic cynomolgus monkeys by using lentiviral infection of monkey oocytes. These TG monkeys showed an increased frequency of repetitive circular locomotion, increase anxiety, reduced social interaction and relatively weak cognitive phenotypes. Overall, we found no evidence of correlation between the copy number of transgenes and the extent of behavioural abnormalities, presumably owing to the low sample number with each copy number, and the possibility of nonspecific effects of gene transfer on behaviours could not be excluded. Importantly, we generated five F<sub>1</sub> TG offspring from one founder TG monkey, confirming the feasibility of germline transmission of lentiviral-based genetic engineering in monkeys. Together with recent progress in applying new gene-editing methods in monkey<sup>29,30</sup>, our findings pave the way for the efficient use of genetically engineered macaque monkeys for studying brain disorders.



**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 31 May; accepted 14 December 2015.**

**Published online 25 January 2016.**

- Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. & Bird, A. P. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**, 499–507 (1989).
- Nan, X., Campoy, F. J. & Bird, A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88**, 471–481 (1997).
- Young, J. I. et al. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc. Natl Acad. Sci. USA* **102**, 17551–17558 (2005).
- Cheng, T. L. et al. MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex. *Dev. Cell* **28**, 547–560 (2014).
- Amir, R. E. et al. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genet.* **23**, 185–188 (1999).
- Ramocki, M. B. et al. Autism and other neuropsychiatric symptoms are prevalent in individuals with MeCP2 duplication syndrome. *Ann. Neurol.* **66**, 771–782 (2009).
- Samaco, R. C. et al. Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2 duplication syndrome. *Nature Genet.* **44**, 206–211 (2012).
- Collins, A. L. et al. Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* **13**, 2679–2689 (2004).
- Kalin, N. H. & Shelton, S. E. Defensive behaviors in infant rhesus monkeys: environmental cues and neurochemical regulation. *Science* **243**, 1718–1721 (1989).
- Nakagawa, T. et al. Generation of lentiviral transgenic rats expressing glutamate receptor interacting protein 1 (GRIP1) in brain, spinal cord and testis. *J. Neurosci. Methods* **152**, 1–9 (2006).
- Frye, R. E., Melnyk, S. & Macfabe, D. F. Unique acyl-carnitine profiles are potential biomarkers for acquired mitochondrial disease in autism spectrum disorder. *Transl. Psychiatry* **3**, e220 (2013).
- Yan, G. et al. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nature Biotechnol.* **29**, 1019–1023 (2011).
- Higashino, A. et al. Whole-genome sequencing and analysis of the Malaysian cynomolgus macaque (*Macaca fascicularis*) genome. *Genome Biol.* **13**, R58 (2012).
- Huh, J. W. et al. Large-scale transcriptome sequencing and gene analyses in the crab-eating macaque (*Macaca fascicularis*) for biomedical research. *BMC Genomics* **13**, 163 (2012).
- Chahrouh, M. et al. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320**, 1224–1229 (2008).
- Ramiro, L. S., Madrid, B. J. & Brown, D. W. Adverse childhood experiences (ACE) and health-risk behaviors among adults in a developing country setting. *Child Abuse Negl.* **34**, 842–855 (2010).
- Drago, L. & Thierry, B. Effects of six-day maternal separation on tonkean macaque infants. *Primates* **41**, 137–145 (2000).
- Suomi, S. J. Early determinants of behaviour: Evidence from primate studies. *Br. Med. Bull.* **53**, 170–184 (1997).
- Feng, X. L. et al. Maternal separation produces lasting changes in cortisol and behavior in rhesus monkeys. *Proc. Natl Acad. Sci. USA* **108**, 14312–14317 (2012).
- Bauman, M. D., Lavenex, P., Mason, W. A., Capitanio, J. P. & Amaral, D. G. The development of social behavior following neonatal amygdala lesions in rhesus monkeys. *J. Cogn. Neurosci.* **16**, 1388–1411 (2004).
- Emery, N. J. et al. The effects of bilateral lesions of the amygdala on dyadic social interactions in rhesus monkeys (*Macaca mulatta*). *Behav. Neurosci.* **115**, 515–544 (2001).
- Golub, M. S., Hogrefe, C. E. & Germann, S. L. Iron deprivation during fetal development changes the behavior of juvenile rhesus monkeys. *J. Nutr.* **137**, 979–984 (2007).
- Makori, N., Watson, R. E., Hogrefe, C. E., Lalayeva, N. & Oneda, S. Object discrimination and reversal learning in infant and juvenile non-human primates in a non-clinical laboratory. *J. Med. Primatol.* **42**, 147–157 (2013).
- Sackett, G., Ruppenthal, G., Hewitson, L., Simerly, C. & Schatten, G. neonatal behavior and infant cognitive development in rhesus macaques produced by assisted reproductive technologies. *Dev. Psychobiol.* **48**, 243–265 (2006).
- Harlow, H. F. The development of learning in the rhesus monkey. *Am. Sci.* **47**, 458–479 (1959).
- Ha, J. C., Mandell, D. J. & Gray, J. Two-item discrimination and Hamilton search learning in infant pigtailed macaque monkeys. *Behav. Processes* **86**, 1–6 (2011).
- Sasaki, E. et al. Generation of transgenic non-human primates with germline transmission. *Nature* **459**, 523–527 (2009).
- Moran, S. et al. Germline transmission in transgenic Huntington's disease monkeys. *Theriogenology* **84**, 277–285 (2015).
- Liu, H. et al. TALEN-mediated gene mutagenesis in rhesus and cynomolgus monkeys. *Cell Stem Cell* **14**, 323–328 (2014).
- Liu, Z. et al. Generation of a monkey with MECP2 mutations by TALEN-based gene targeting. *Neurosci. Bull.* **30**, 381–386 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank M.-m. Poo for comments on the manuscript, Y.-Z. Li, Y.-Z. Lu, F. Liu and X. Zhou for maintaining monkey colony, W. Lu, X.-H. Guo and Y. F. Zhou of Fudan Children's Hospital for assistance in mass spectrometry and electroencephalogram, D. Chen and B. Zhang of Novel Bioinformatics for RNA-seq data analysis, and C.-H. Li of Shanghai GenesKies Company for analysis of genomic integration sites. This work was supported by CAS Strategic Priority Research Program (XDB02050400), the MoST 973 Program (2011CBA00400), NSFC grants (91432111, 91232712 and 81527901), National Key Technology R&D Program of China 2014BAI03B00, Shanghai City Committee of Science and Technology Project 14140900100.

**Author Contributions** Z.Q. and Q.S. conceived and supervised the project. T.-L.C. constructed the lentiviral constructs. Q.S. and Z.L. performed the cynomolgus oocytes preparation and injection. Y.-J.C., Y.W., C.-C.Z., Y.-H.N. and Z.L. contributed to monkey reproductive experiments. Y.-F.Z. performed PCR-based genotyping experiments. Z.-F.C., W.-J.B., X.-D.Z. and X.Y. performed immunohistochemistry and AccuCopy experiments. C.C., B.L., X.S. and Z.-Q.X. performed western blot experiments. X.L. and J.-J.W. performed behavioural analysis. J.-T.Z. and N.G. performed WGTA tests. W.-H.Z. and X.X. contributed to metabolic measurements and behavioural analysis. T.-L.C. and X.L. performed genomic integration sites analysis based on deep-sequencing. J.X., L.Z. and F.Z. helped with identification of genomic integration sites of transgenes. Z.Q. wrote the manuscript.

**Author Information** The raw sequence and processed data have been submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE57974. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.Q. (zqiu@ion.ac.cn) or Q.S. (qsun@ion.ac.cn).

## METHODS

**Animal ethics statement.** The use and care of animals complied with the guideline of the Biomedical Research Ethics Committee at the Shanghai Institutes for Biological Science (CAS), which approved the application entitled 'Reproductive physiology of cynomolgus monkey and establishment transgenic monkey' (#ER-SIBS-221106P).

**Collection of oocytes, gene delivery and embryo construction.** Laparoscopy was used for oocyte collection. Oocytes were aspirated from follicles 2–8 mm in diameter, about 32–36 h after hCG stimulation<sup>31</sup>. The collected oocytes were cultured in the pre-equilibrated maturation medium<sup>32</sup>. Metaphase II arrested oocytes were selected for perivitelline space injection<sup>32</sup> of lentiviruses and ICSI. The lentivirus concentration for injection was  $1 \times 10^{10}$  viral genome (vg) per ml. After microinjection, the oocytes were cultured in the maturation medium at 37°C (in 5% CO<sub>2</sub>) for about 1 h, until fertilization by ICSI. Monkey semen was collected by penile electro-ejaculation. For ICSI, a single sperm was immobilized and aspirated with the tail first. A single oocyte was fixed by the holding pipette, and the injection pipette was pushed through the zona pellucida and subsequently through the oolemma to release the spermatozoon<sup>32</sup>. After ICSI, the oocytes were cultured in pre-equilibrated Hamster Embryo Culture Medium 9 (HECM-9) at 37°C (in 5% CO<sub>2</sub>) until the next morning<sup>33,34</sup>.

**Selection of surrogate females and embryo transfer.** Menstrual cycles of females were recorded daily. To synchronize the developmental stage of embryos with the recipient, monkeys were chosen for tubal embryo transfer at 0–3 days after ovulation, and a stigma or a new corpus luteum on the ovary could be observed by laparoscopy. About 2–3 pronuclear-stage embryos were selected for tubal transfer to each surrogate female<sup>31</sup>.

**Genotyping PCR.** Hair-root samples collected from newborn monkey pups were used to extract DNA. Samples were digested by proteinase K overnight at 65°C and precipitated for DNA and PCR with specific primers again GFP and mCherry were used for initial genotyping analysis as follows: mCherry-R: 5'-TGCTTGATCTCGCCCTTCAG-3', mCherry-F: 5'-GCCATCATCAAGGAGTTTCATGC-3'; GFP-F: 5'-AAGTTCATCTGCACCACCG-3', GFP-R: 5'-TCCTTGAAGAAGATGGTGGC-3'.

**Southern blot analysis.** A total of 15 µg of genomic DNA was prepared and digested with BamHI and EcoRI, which released transgenes. Genomic DNAs were separated with 1% agarose gel and transferred to Nippon N+ membrane (GE). DNA probes from hMECP2-2a-GFP was prepared using ready-to-go DNA label kit (279240D-20, GE Life Sciences). <sup>32</sup>P-labelled probes were hybridized with blots of genomic DNAs and exposed to phosphor-imager after extensively washing.

**Euthanasia procedure for sick and aborted newborn monkeys.** Decisions of whether euthanasia procedures would be carried out for sick or aborted newborn monkeys are made by veterinarians, after consulting with principal investigators and followed the approved protocol (#ER-SIBS-221106P). Aborted or sick MECP2 TG and WT monkeys were deeply anaesthetized with ketamine hydrochloride (5–10 mg kg<sup>-1</sup>) to avoid possible pain and then perfused with 0.9% saline with 2–4% paraformaldehyde (PFA) for further immunohistochemistry experiments. The procedure is approved by the Biomedical Research Ethics Committee at the Shanghai Institutes for Biological Science (CAS), described in the protocol entitled 'Reproductive physiology of cynomolgus monkey and establishment transgenic monkey' (#ER-SIBS-221106P).

**Immunohistochemistry of transgenic monkey tissues.** After perfusion, the hemispheres of the brain were dissected, cut in to small blocks, fixed with 4% PFA in phosphate buffer, and equilibrated in 30% sucrose. Fixed and equilibrated brain tissue blocks were cut into 30-µm cortical sections with a Microm HM525 cryostat. Sections were washed for 5 min in PBS containing 5% bovine serum albumin (BSA) and 0.3% Triton X-100, and incubated with primary antibodies (in PBS with 3% BSA and 0.3% Triton X-100) overnight at 4°C and subsequently with corresponding secondary antibodies (Alexa-Fluor-conjugated, Invitrogen, at 1:1,000). DAPI was used to label the nuclei and sections were mounted with 75% glycerol. Other antibodies used: HA antibody (Covance, MMS-101R), NeuN antibody (Millipore, MAB377), MeCP2 antibody (Cell Signaling, 3456S) and GFP antibody (Abcam, ab6673).

**Measurement of MECP2 transgene copy numbers using AccuCopy assay.** Four sets of primers targeted to MECP2 were designed. One set (mecp2\_1) was a cross-intron primer targeted to transgenic cDNA fragments representing the copy number of transgenic DNA; the second (mecp2\_2) was targeted to one exon of transgenic cDNA fragments representing the total MECP2 copy number; and the other two primer sets (mecp2IN\_1 and mecp2IN\_2) were targeted to introns of monkey MECP2 gene representing the endogenous MECP2 copy number. Two sets of EGFP primers (EGFP\_1 and EGFP\_2) were designed to verify the copy number of the transgene, and one set of mCherry primers was designed as negative control. The copy number of these DNA fragments was measured using custom-designed

Multiplex AccuCopy™ Kit (Genesies Biotechnologies, CN0105). The copy number of these target DNA fragments was measured using custom-designed Multiplex AccuCopy kit (Genesies Biotechnologies, CN0105). For each DNA fragment amplified, a piece of synthesized competitive double-stranded DNA of known concentration and with insertions or deletions of a few base pairs was added to the PCR reaction mix. Each PCR reaction was carried out by mixing the synthesized competitive double-stranded DNAs for target and reference genes (POP1, RPP14 and POLR2A) together with a defined amount of sample DNAs. A multiplex competitive PCR was then performed to simultaneously amplify all reference and target genes from both sample and competitive DNAs using multiple fluorescence-labelled primer pairs. In brief, the 20-µl PCR reaction for each sample contained 1 × AccuCopy PCR Master Mix, 1 × Fluorescence Primer Mix, 1 × Competitive DNA mix and ~10 ng sample DNA. The PCR program used was: 95°C for 10 min; 11 cycles of 94°C for 20 s, 65°C–0.5°C/cycle 40 s, 72°C for 1.5 min; 24 cycles of 94°C for 20 s, 59°C for 30 s, 72°C for 1.5 min; 60°C for 60 min. PCR products were diluted 20-fold before loaded on ABI3730XL sequencer (Applied Biosystems) to separate amplicons of different sizes by capillary electrophoresis. Raw data were analysed using GeneMapper4.0, and the peak ratios of sample DNA to competitive DNA (S/C ratio) for all target and reference fragments were exported to Excel. The S/C ratio of each target fragment was first normalized to the S/C ratio of the reference genes, and then further normalized to the median copy number of the entire data set. The final normalized ratio was averaged for each MECP2 primer and EGFP primer, and the similarity between the two ratio further confirmed the copy number of the transgene.

Primer sequences: mecp2\_1-for 5'-CGCTCTGCTGGGAAGTATGATG-3', mecp2\_1-rev 5'-GGGATGTGTCGCTACCTTTTC-3'; mecp2\_2-for 5'-AAGCCC AAATCTCCCAAAGTC-3', mecp2\_2-rev 5'-TTCCAGGACTTTTCTC CAGGAC-3'; mecp2IN\_1-for 5'-GCAAGTTTGGCTGAAGGAGAA-3', mecp2IN\_1-rev 5'-GAGCACACCCACAGCAGTAA-3'; mecp2IN\_2-for 5'-TTGGGCTTGAAGTCCAGACCTC-3', mecp2IN\_2-rev 5'-GGTGGTGCCAC TGACTGAGAAG-3'; EGFP\_1-for 5'-TACGGCAAGCTGACCCTGAAGT-3', EGFP\_1-rev 5'-CTGCTTCATGTGGTCCGGGTAG-3'; EGFP\_2-for 5'-CTGCC CGCAACCACTACCTG-3', EGFP\_2-rev 5'-GAATCCAGCAGGACCAT GTGA-3'; POP1-for (internal control) 5'-AGAAATGCTTGGGCTGTTACG-3'; POP1-rev (internal control) 5'-CACAGCTGCCTGCTCTCAGAAG-3'; RPP14-for (internal control) 5'-TTTTTTGGTGCCTGCATGTTTG-3', RPP14-rev (internal control) 5'-AGCGATGGCTGGAAATTAGTGG-3'; POLR2A-for 5'-GGTGAAGTACGACGCGACTGTG-3', POLR2A-rev 5'-AGGTTTCAGCGTAGCCAGGTTTC-3'.

**Preparation of lentivirus.** Lentiviruses were produced by standard protocols and provided at a titre of  $10^{10}$  vg ml<sup>-1</sup> by the Shanghai SBO Medical Biotechnology Co. Ltd.

**Identification of genomic integration sites of transgenes.** A total of 2 µg genomic DNA was used to construct a DNA library for each case<sup>35–38</sup>. Sequencing linkers were further added onto genomic segments (length around 500–700 base pairs (bp)) (Extended Data Fig. 1b). After end repairing and 3' A-adding, the fragmented DNAs were ligated with Y-shape adaptor. Amplification was performed with the adaptor primers. Asymmetry-primer PCR (APP) was used to enrich the viral integration sites in each library. The APP method includes two PCR systems. The first PCR system includes only LTR specific primer. After 12 cycles of linear amplification, adaptor specific primer was added in the PCR system followed by 12 cycles of exponential amplification. PCR products were purified using 0.7 × AMPure beads (Beckman, A63882). The second PCR system uses a pair of primers nest the primers in the first PCR system. After 12 cycles of linear amplification and 15 cycles of exponential amplification, the PCR products of 500–700 bp in size were isolated by agarose gel electrophoresis before being used to construct libraries with Illumina paired-end adaptors according to the manufacturer protocol and sequenced by Illumina MiSeq V3 (2 × 300 base paired ends). Only the paired-end reads showing the fusions of viral sequences and the cynomolgus (*Macaca fascicularis*) genome segments were selected, in which two mismatches were allowed. The reads showing the same integration position were merged and treated as a unique integration site. Experiments were repeated three times independently with different sequencing linkers. Determination of insertion sites is under the following criteria: (1) total insert numbers are greater than 100 times after three experiments; (2) being detected at least twice after three experiments. Cynomolgus monkey genome is used in the following database: <http://www.ncbi.nlm.nih.gov/genome/?term=crab+eating+monkey>. Target sequences containing LTR of transgene cassettes and genomic segments flanking the transgenes were analysed (Supplementary Tables 2 and 6).

Primer sequences: LTR\_1: 5'-CTTGCCCTTGAGTGCTTCAAGTAGT-3'; LTR\_2: 5'-TGCCCGTCTGTGTGTAAGTCT-3'; YP\_1: 5'-GGATAGCGACGCA CGGAAGTCT-3'; YP\_2: 5'-CTCCATCTCATCTGCGTGTG-3'.

**Western blotting.** Monkey brain tissues were homogenized in RIPA buffer (containing 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% Triton X-100, 0.1% SDS, 1% sodium deoxycholate, protease inhibitor cocktail and phosphatase inhibitor cocktail) on ice and then centrifuged at 1,000g for 10 min at 4°C. The supernatant was stored at -80°C until use. Protein concentration was measured with BCA method. Approximately 30 µg protein of each sample was loaded in on 10% SDS-PAGE and run at 120 V constant voltage. A constant current of 0.36 mA was used for transblotting. Blots were probed with primary antibodies (1:1,000) overnight at 4°C. After washing three times, blots were then incubated with goat anti-rabbit secondary antibody (1:3,000) at room temperature for 2 h. Chemiluminescence was used to visualize protein bands. Antibodies used: HA antibody (Abcam, ab9110), MeCP2 antibody (Cell Signaling, 3456S) and GFP antibody (Invitrogen, A11122).

**Mass spectrometry analysis.** Fresh whole blood from upper arms of monkeys was taken by a professional veterinarian in the morning before feeding. Whole blood (200–400 µl) was dropped onto filter paper immediately. After air drying, filter papers were store at -20°C before mass spectrometry analysis. API2000 from AB SCIEX were used for analysing fatty acid and amino acids. Data were obtained from three rounds of blood collections independently.

**Behaviour observation and analysis.** Behaviour observation and analysis were performed by two independent trained observers, with demonstrated inter-observer reliability of at least 80%. All observers were blinded to the genotypes of the monkeys. The dimension of cages using for living and behaviour monitoring is  $1.5 \times 1 \times 1.1$  m.

**Analysis of locomotive behaviour.** Monkeys individually, were observed alone in an observation cage ( $1.5 \times 1 \times 1.1$  m) after they had been accustomed to community living following weaning. The observation cage was similar to their home cage. All locomotion behaviours were video-record without interruption for 20 min each day for 5 days. Data from 5 days were pooled.

**Analysis of social interaction behaviours.** Social behaviours of TG and WT monkeys with familiar and unfamiliar monkeys were studied by examining the interactions of monkey from the same and different home cage, respectively. To study the interaction with familiar monkeys, we housed three groups of monkeys, each consisting three WT and two TG monkey of the same age, in three separate cages for 6 months before the observation (at about 1.5 years old). In this analysis, the observer followed the time each monkey spent sitting together with another monkey for a duration of 1 h each day for 5 consecutive days. We defined that two monkeys sat together by obvious interactions between the two for more than 3 s, during which the monkeys may exhibit touching and grooming behaviours or lean against each other. To study interaction with unfamiliar monkeys, we regrouped the females from same cohorts after the above observation for another 8 months in four separate cages (see Supplementary Table 4a, b). (Males were kept together separately owing to their proximity to sexual maturity, thus not used for observation). For each observation of social interaction, we paired two monkeys from different group and observation was made in the same manner as that described above for the interaction between familiar monkeys.

**Analysis of social interaction behaviours of F<sub>1</sub> generations.** To study the interaction with F<sub>1</sub> TG monkeys, we housed two groups of monkeys (group info see Supplementary Table 7), each consisting of three WT and two TG F<sub>1</sub> monkeys of similar age (at 10–11 months old), in two separate cages before the observation. In this analysis, the observer followed the time each monkey spent sitting together with another monkey for a duration of 1 h each day for 5 consecutive days.

**Analysis of TAD behaviours.** The TAD behavioural model was used to assay the monkey's response to human gaze (Extended Data Fig. 5a). In each session of observation, an individual monkey from either the transgenic or WT group was placed in an observation cage ( $1.5 \times 1 \times 1.1$  m), and allowed to adapt to the cage alone for 9 min. An observer then sat in front of the cage at a distance of 2 m, showing the face profile to the monkey without eye contact for 9 min ('non-gaze period'). This was followed by the relaxation period (3 min) without the human presence, and the 'gaze period' (9 min) in which the observer sat in front of the cage and gazed at monkey with a neutral face. Behaviour and vocalizations were recorded on videotape<sup>39–41</sup>.

**WGTA.** WGTA tests were performed on 8 TG and 6 WT monkeys at the age of 1.5 years, in accordance to WGTA protocol<sup>25,26</sup>, by trained technicians. The WGTA apparatus includes a testing box that for observing subject's activity, a presentation board with food wells for reward placing, a trial door and an access door connected by pulley cord to separate the subject and presentation board, and a camera for recording. All tests were carried out in a quiet and standard lighted room.

**Black/white test.** This test includes three stages: adaptation, discrimination and reversal. For the adaptation step, each monkey was tested for the ability to take the food reward on the presentation board that was placed by experimenter. Before the adaptation step, the monkey needs to pass several pre-test steps: the reward was placed in front of the food well, in the food well, in the food well next to the

adaptation block, and in the food well with half covered by the adaptation block. Finally for adaptation step, the monkey had to take the food in the food well with the block covered completely. Each monkey received a maximum of 25 trials per day, and was considered to be passed when showing correct responses on 23 out of 25 trials. During the discrimination step, each monkey needed to choose the only reward in the food well that was covered by either a black or white block with an empty well covered the opposite colour. The same monkey was always rewarded with either black or white but with random location, with assignment of monkeys by the Gellerman order. Each monkey received 25 trials per day and was considered to be passed when showing correct responses on 23 out of 25 trials. For the reversal step, the procedure was the same as discrimination step, except that the monkey was rewarded black if white was rewarded during the discrimination step, and vice versa.

**Hamilton search test.** This test includes four steps: adaptation, Hamilton search, Hamilton search set-breaking, and Hamilton search forced set-breaking. The adaptation step was similar to that for black/white test, with the same criterion for passing. For the Hamilton search step, four little boxes that represented the different positions from experimenter's left to right on the presentation board were used for testing. The only reward was randomly placed in one of the four closed boxes in each trial. Monkey was allowed to find the reward from these four closed boxes. One trial was terminated when the monkey open the correct box. Each monkey performed 25 trials per day for 5 consecutive days. For the Hamilton search set-breaking step, the box that was the least preferred was first determined from the above step, and was always rewarded when chosen by the monkey. One trial was terminated when the subject open the correct box. Each monkey performed 25 trials per day for 5 days. For the Hamilton search forced set-breaking step, the procedure was the same as the set-breaking test, except that the monkey was allowed to make only one choice for finding the reward that placed in the least preferred box. The monkey was scored for the rate of correct choice over 25 trials each day for 5 consecutive days.

**Learning set test.** The monkeys were tested for the ability to distinguish 240 pairs of toys. The toys in each pair were labelled A or B to cover the two food wells, one of which had food. For each monkey, either A or B was always rewarded. Each pair of toys was presented for 6 trials and 6 pairs were tested each day. Six different pairs were used for different days, with the test lasting 8 weeks until all 240 pairs were used. The monkey was scored for the rate of correct choice, averaged over 180 trials (5 days).

**Transcriptome analysis for transgenic monkey.** Total RNA was extracted from three independent pieces of cortical tissues from brains of T05, T07, T09 and T14 and four WT monkeys by Trizol reagent (Invitrogen) separately. The RNA quality was checked by Bioanalyzer 2200 (Agilent) and kept at -80°C. The RNA with RIN (RNA integrity number) > 8.0 is acceptable for cDNA library construction. RNA-seq and bioinformatic data analysis were performed by Shanghai Novellbio Ltd.

**Library construction and RNA-seq.** The cDNA libraries for single-end sequencing were prepared using Ion Total RNA-Seq Kit v2.0 (Life Technologies) according to the manufacturer's instructions. The cDNA libraries were then processed for the proton sequencing process according to the commercially available protocols. Samples were diluted and mixed, the mixture was processed on a OneTouch 2 instrument (Life Technologies) and enriched on a OneTouch 2 ES station (Life Technologies) for preparing the template-positive Ion PI Ion Sphere Particles (Life Technologies) according to Ion PI Template OT2 200 Kit v2.0 (Life Technologies). After enrichment, the mixed template-positive Ion PI Ion Sphere Particles of samples was loaded on to 1 P1v2 Proton Chip (Life Technologies) and sequenced on Proton Sequencers according to Ion PI Sequencing 200 Kit v2.0 (Life Technologies).

**RNA-seq mapping.** Before read mapping, clean reads were obtained from the raw reads by removing the adaptor sequences, reads with >5% ambiguous bases (noted as N) and low-quality reads containing more than 20% of bases with qualities of <13. The clean reads were then aligned to crab eating macaque genome (version: Mfa5.0) using the MapSplice program (v2.1.6). In alignment, preliminary experiments were performed to optimize the alignment parameters (-s 22 -p 15 -ins 6-del 6-non-canonical) to provide the largest information on the AS events<sup>42</sup>.

Dif-Gene-Find er-t. We applied DEseq algorithm to filter the differentially expressed genes, after the significant analysis and false discovery rate (FDR) analysis under the following criteria: (1) fold change > 1.5 or < 0.667; (2) FDR < 0.05 (ref. 43). A Volcano plot was drawn by P value based on the differential gene analysis, and the colour was determined by the filtering criteria (red,  $\log_{10}(P \text{ value}) > 1.5$ ; blue,  $\log_{10}(P \text{ value}) < 1.5$ ; black,  $\log_2(\text{FC}(\text{TG}/\text{WT})) < \pm 0.5$ ).

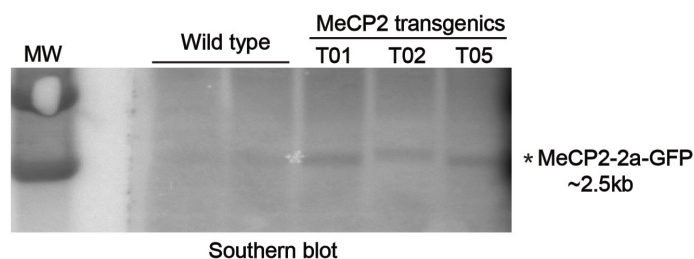
**Generation of F<sub>1</sub> offspring of transgenic monkeys.** The F<sub>1</sub> offspring was generated by ICSI using sperms obtained from testicular tissue xenografts of the T07 monkey. The method of testicular xenografting greatly shortened the time required for sexual maturation of TG monkey<sup>44</sup>.



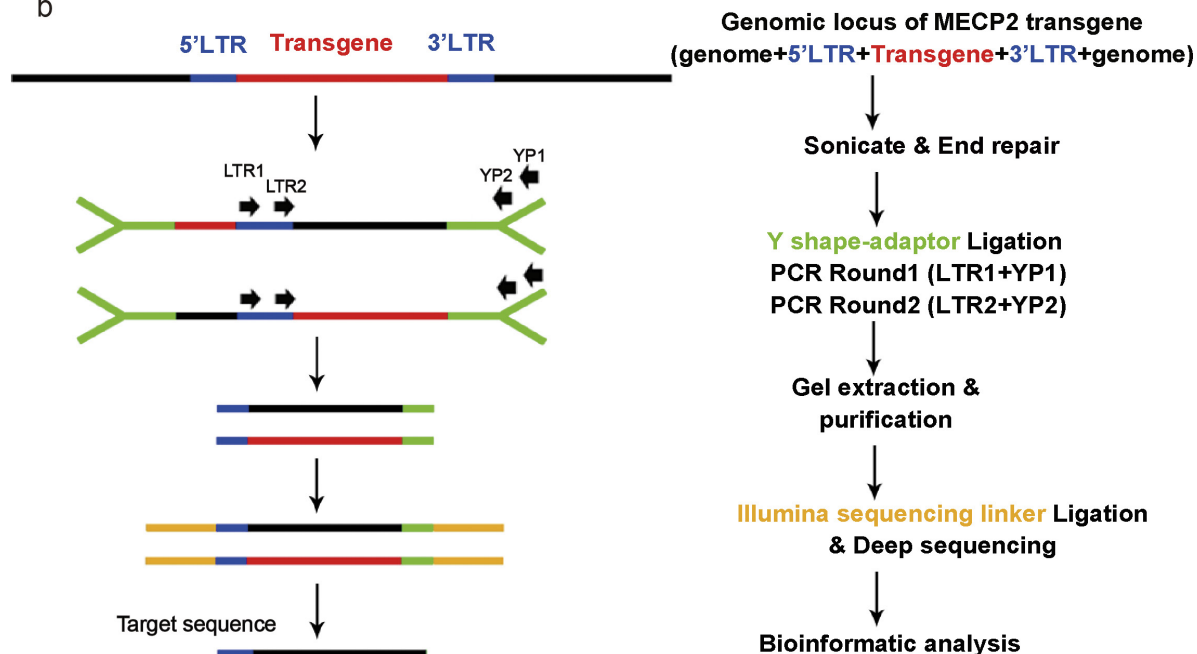
31. Sun, Q. *et al.* Efficient reproduction of cynomolgus monkey using pronuclear embryo transfer technique. *Proc. Natl Acad. Sci. USA* **105**, 12956–12960 (2008).
32. Chan, A. W. & Yang, S. H. Generation of transgenic monkeys with human inherited genetic disease. *Methods* **49**, 78–84 (2009).
33. Zheng, P., Bavister, B. D. & Ji, W. Z. Energy substrate requirement for in vitro maturation of oocytes from unstimulated adult rhesus monkeys. *Mol. Reprod. Dev.* **58**, 348–355 (2001).
34. Zheng, P., Wang, H., Bavister, B. D. & Ji, W. Maturation of rhesus monkey oocytes in chemically defined culture media and their functional assessment by IVF and embryo development. *Hum. Reprod.* **16**, 300–305 (2001).
35. Janovitz, T. *et al.* High-throughput sequencing reveals principles of adeno-associated virus serotype 2 integration. *J. Virol.* **87**, 8559–8568 (2013).
36. Cohn, L. B. *et al.* HIV-1 integration landscape during latent and active infection. *Cell* **160**, 420–432 (2015).
37. Xiao, J. *et al.* Rearrangement structure-independent strategy of CNV breakpoint analysis. *Mol. Genet. Genomics* **289**, 755–763 (2014).
38. Du, R. *et al.* Efficient typing of copy number variations in a segmental duplication-mediated rearrangement hotspot using multiplex competitive amplification. *J. Hum. Genet.* **57**, 545–551 (2012).
39. Kalin, N. H., Shelton, S. E., Davidson, R. J. & Kelley, A. E. The primate amygdala mediates acute fear but not the behavioral and physiological components of anxious temperament. *J. Neurosci.* **21**, 2067–2074 (2001).
40. Kalin, N. H., Shelton, S. E. & Davidson, R. J. The role of the central nucleus of the amygdala in mediating fear and anxiety in the primate. *J. Neurosci.* **24**, 5506–5515 (2004).
41. Rogers, J., Shelton, S. E., Shelledy, W., Garcia, R. & Kalin, N. H. Genetic influences on behavioral inhibition and anxiety in juvenile rhesus macaques. *Genes Brain Behav.* **7**, 463–469 (2008).
42. Wang, K. *et al.* MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* **38**, e178 (2010).
43. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
44. Liu, Z. *et al.* Generation of macaques with sperm derived from juvenile monkey testicular xenografts. *Cell Res.* **26**, 139–142 (2015).



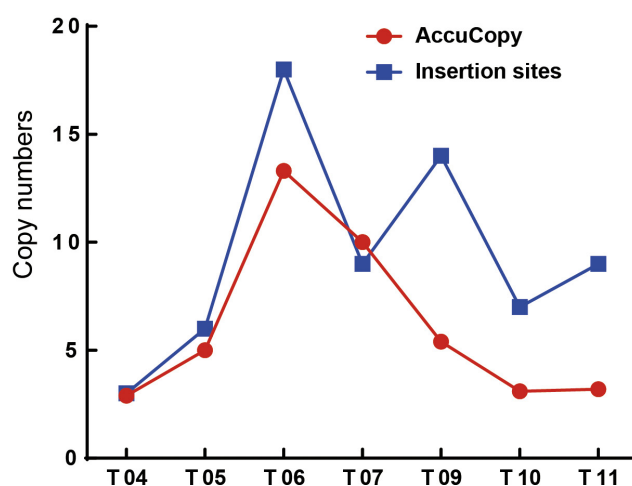
a



b

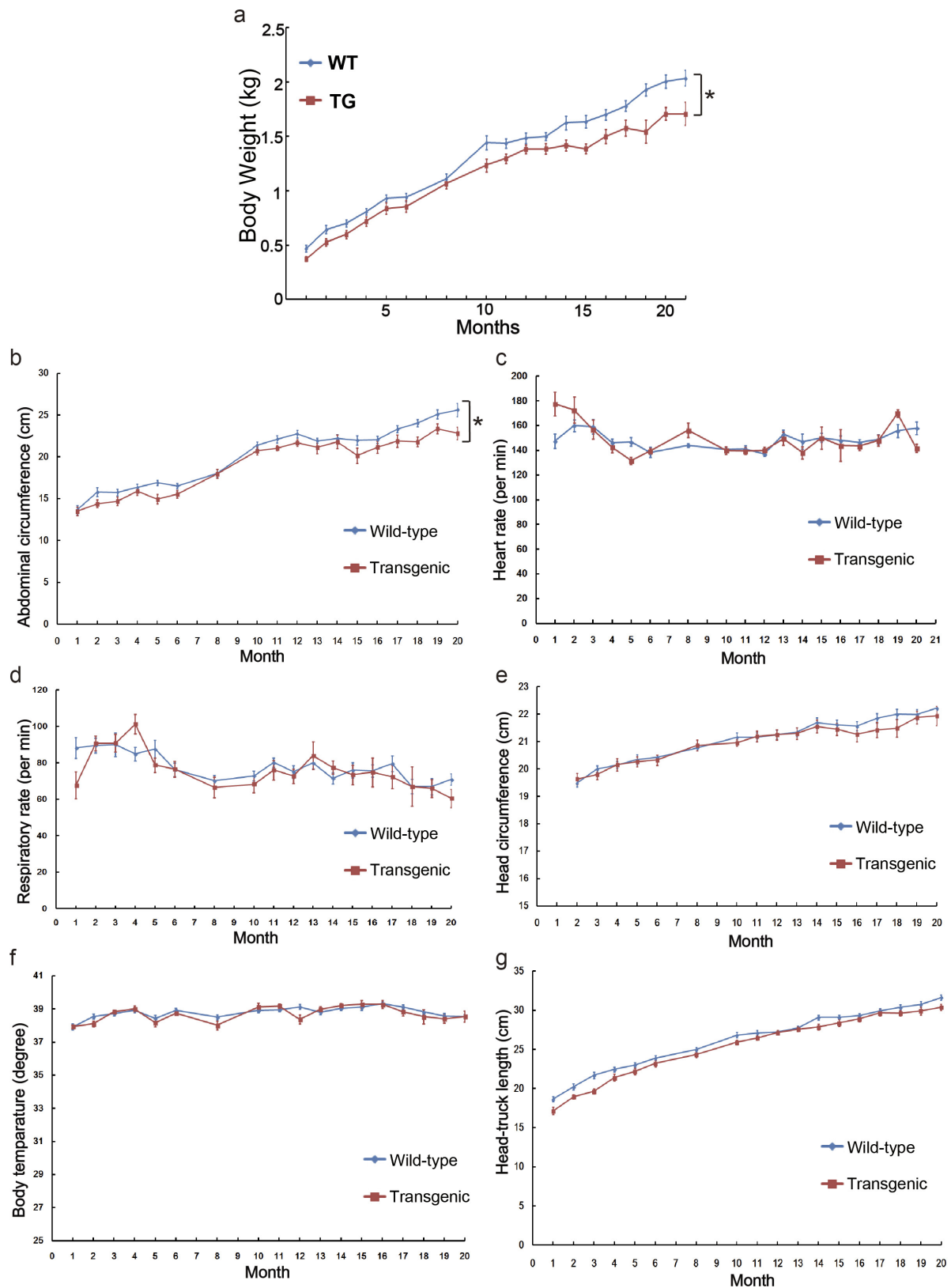


c



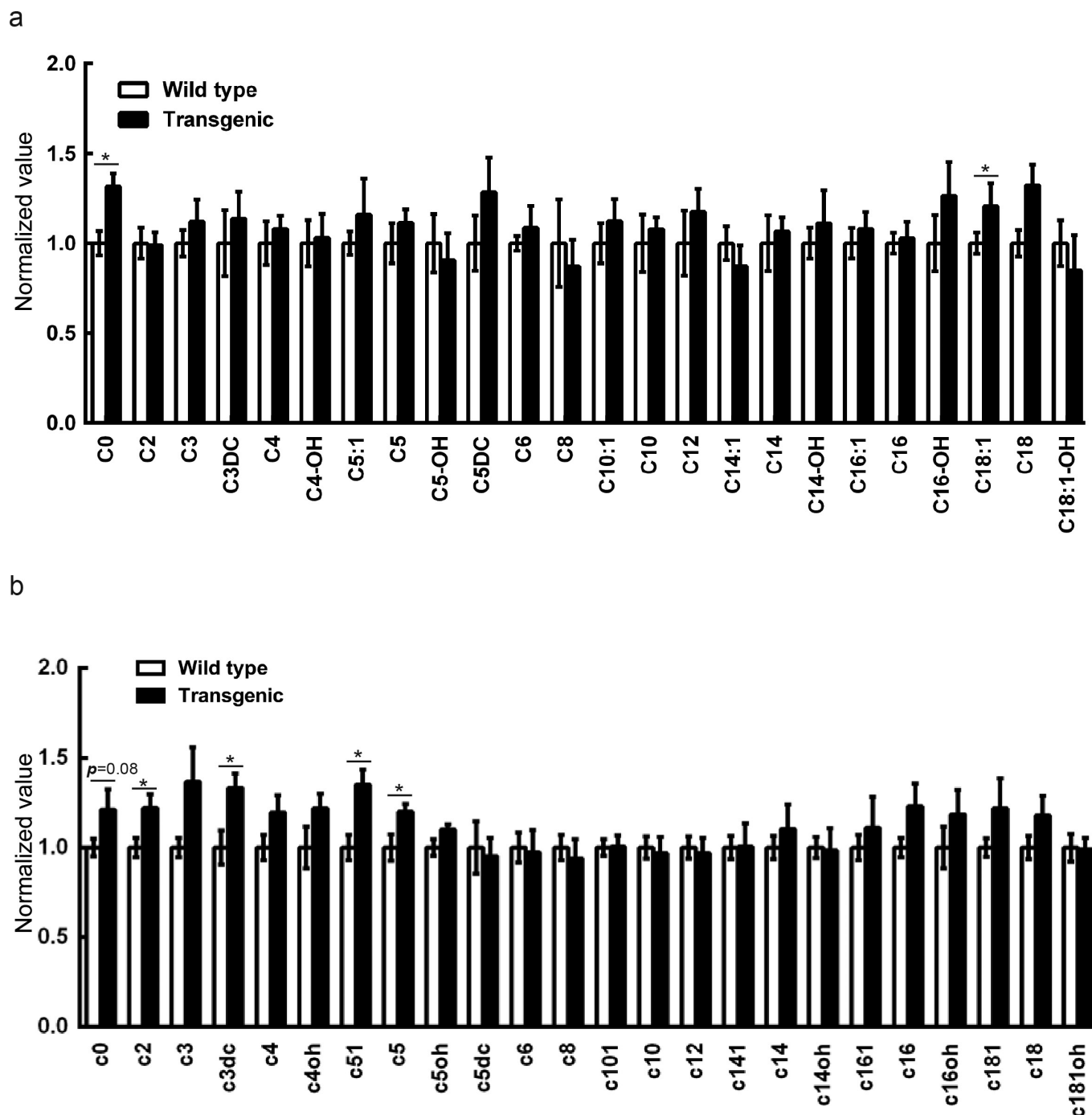
**Extended Data Figure 1 | Analysis of genomic integration sites of transgenes in F<sub>0</sub> TG monkeys.** a, Genomic DNA were extracted from hair roots of WT and TG monkeys, digested and analysed by agarose gel. Radioactive probe labelled by <sup>32</sup>P were prepared against the hMeCP2-2a-GFP transgene. Blots were transferred to membrane and hybridized with the probes. Images were acquired by exposing the blot to a phosphor-imager. Asterisk indicates target band. b, Flowchart of deep-sequencing-based methods for identifying genomic integration sites of lentiviral transgenes. Genomic integration sites of lentiviral transgenes are composed as three parts, endogenous genomic segments (black),

LTRs (blue) and transgenes (red). Genomic DNA was sonicated, end-repaired and ligated to a Y-shaped adaptor, then subjected to two rounds of amplifications by the LTR1 + YP1 and LTR2 + YP2 primer sets. Illumina sequencing linkers were added onto segments and performed paired-end high-throughput sequencing. Target sequences containing LTR and endogenous genomic segments were collected and analysed. c, Comparison of copy numbers obtained from two methods among F<sub>0</sub> TG monkeys. Red denotes copy numbers from AccuCopy (MECP2 and mCherry transgenes); blue denotes LTR insertion sites from deep-sequencing.



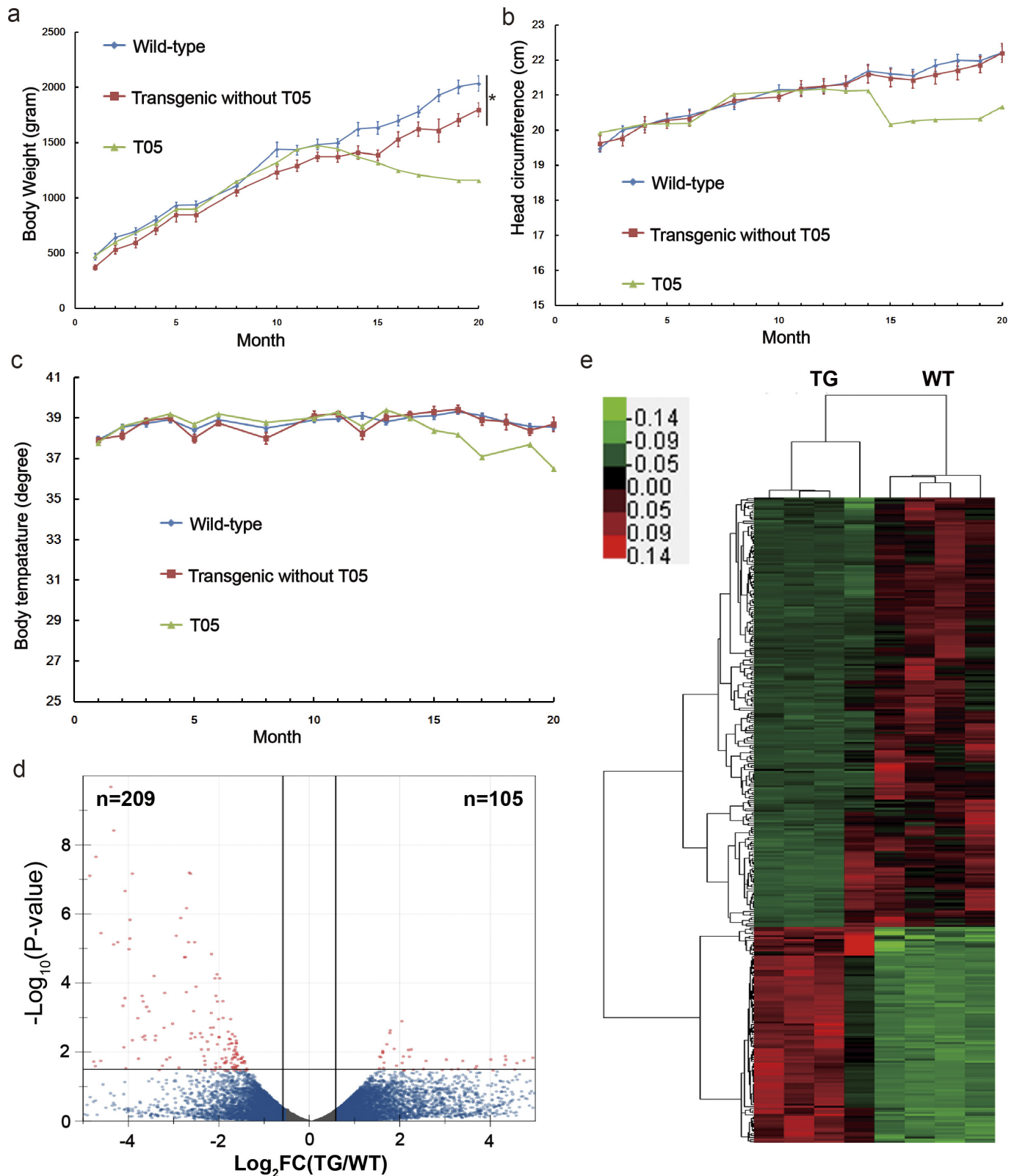
**Extended Data Figure 2 | Physical parameters measured for monkeys.** Developmental changes in the body weight of 8 TG and 18 WT monkeys. **a–g**, Body weight, abdominal circumference, heart rate, respiratory rate,

head circumference, body temperature and head–trunk length were measured for 8 *MECP2* TG and 18 WT monkeys. \* $P < 0.05$  (Mann–Whitney  $U$  test). Error bars denote s.e.m.



**Extended Data Figure 3 | Fatty acid measurements for TG and WT monkeys.** **a**, Blood samples collected at 18 months of age. **b**, Blood samples collected at 36 months of age. The blood levels of different forms of fatty acids were measured by mass spectrometry, with each bar represents

results from three independent samples. C0, total fatty acid contents. All data are normalized to the average values of parallel blood samples from WT monkeys. \* $P < 0.05$  (Student's  $t$ -test). Error bars denote s.e.m.



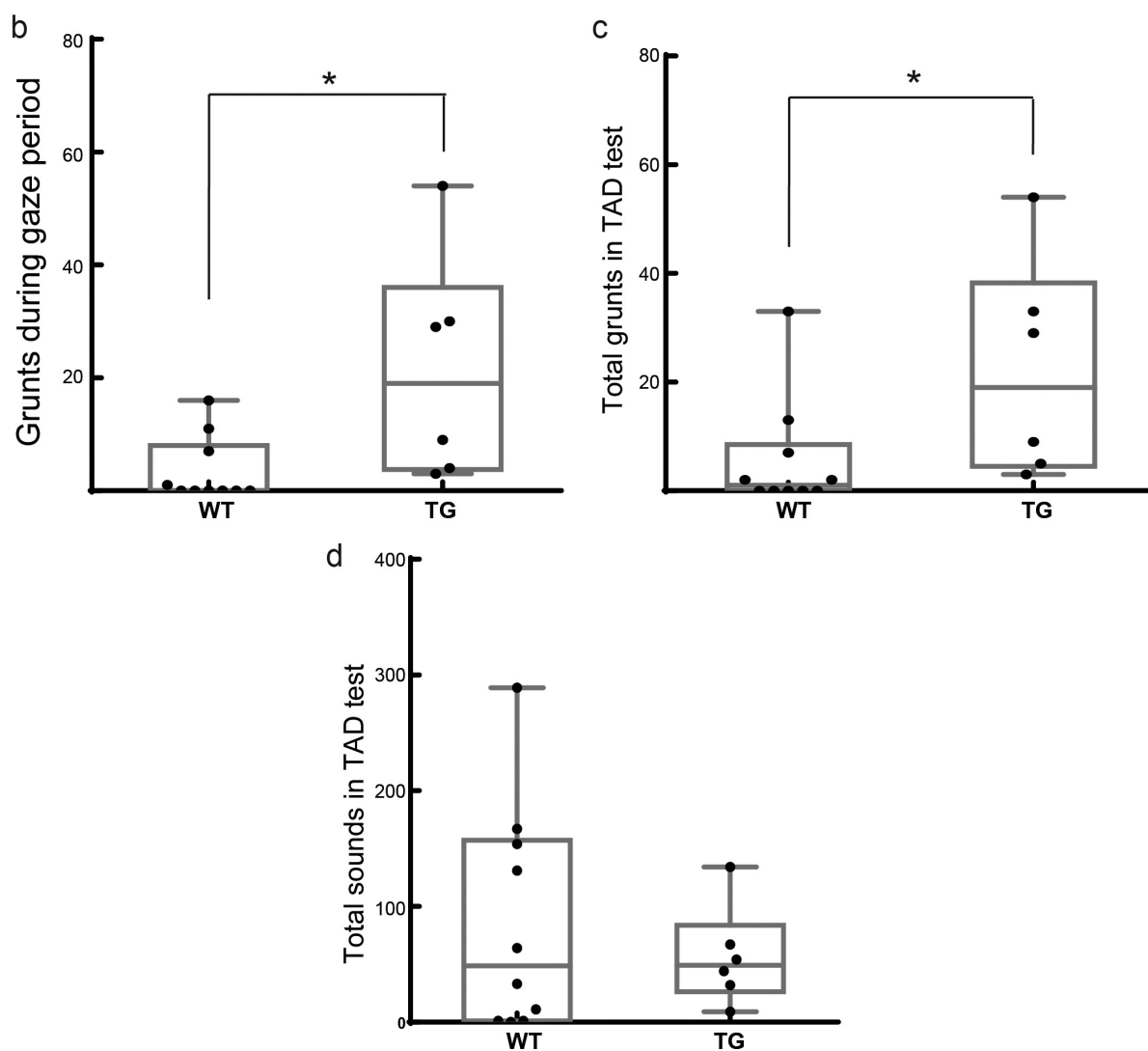
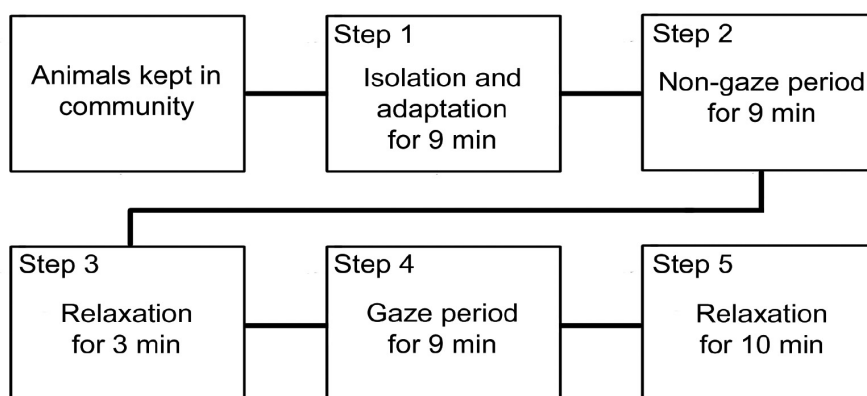
**Extended Data Figure 4 | Physical growth parameters measured for monkey T05 and transcriptome analysis of *MECP2* transgenic monkey.** **a–c**, Body weight, head circumference and body temperature were measured for monkey T05. \* $P < 0.05$  (Mann–Whitney  $U$  test), together with the average data from all other TG and WT monkey monitored. The monkey T05 died at 20 months of age. **d**, Volcano map for alterations in gene expression in the TG monkeys (T14, T05, T07 and T09), as compared to four WT monkeys. Red dots denote genes with a  $>2$ -fold

change (FC) in expression ( $\log_2(FC(TG/WT)) > 1$  or  $< -1$ ). Blue dots denote genes with no significant change in expression ( $P > 0.05$ ). **e**, Heat map representation of the selected genes involved in metabolic pathways and brain development. Gene expression is coded in pseudocolour scale (-0.14 to 0.14). Red denotes higher expression in TG monkeys; green denotes lower expression in TG monkeys, as compared to WT monkeys. Error bars denote s.e.m.



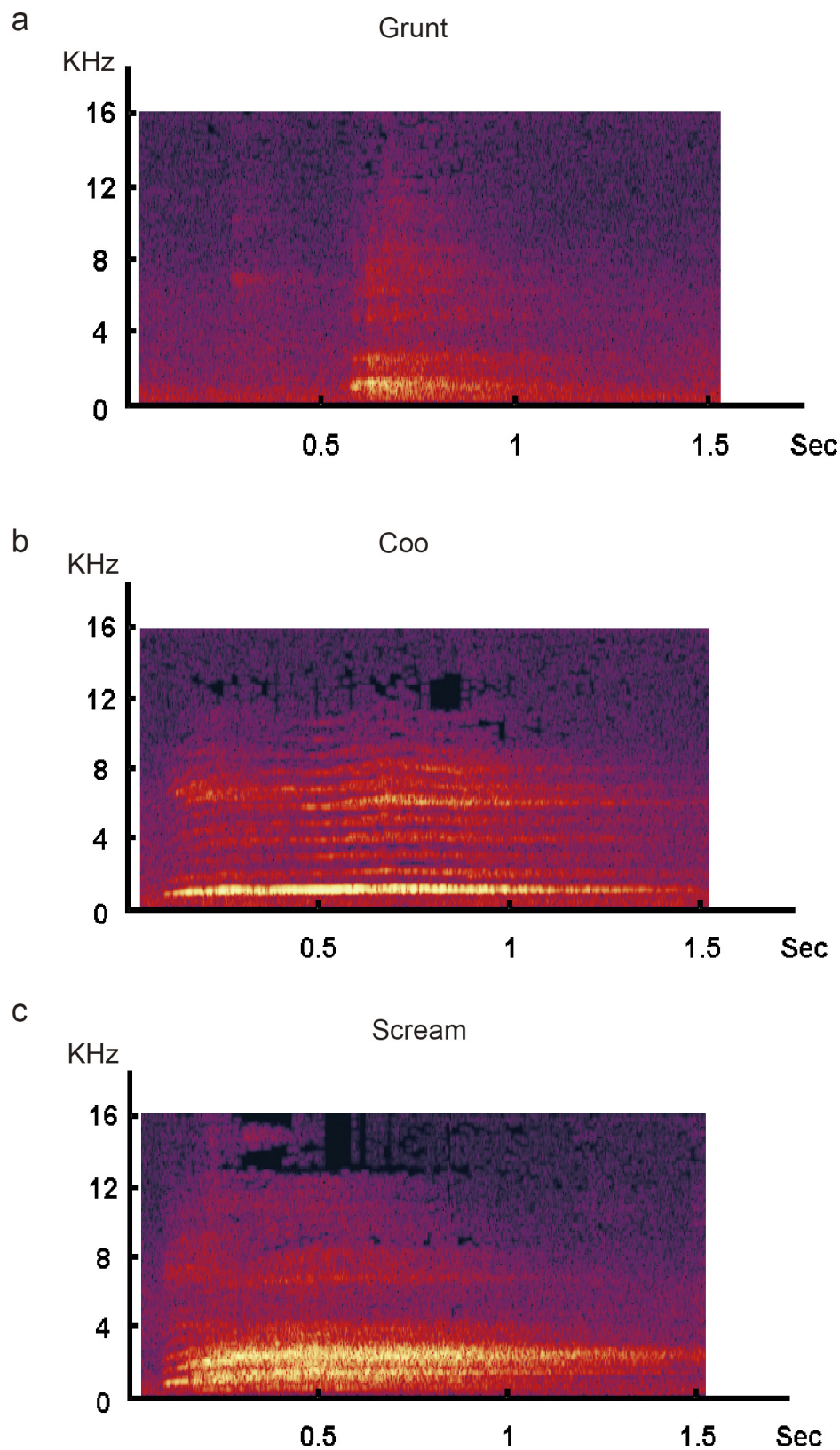
a

## Threat-related Anxiety and Defensive behaviors

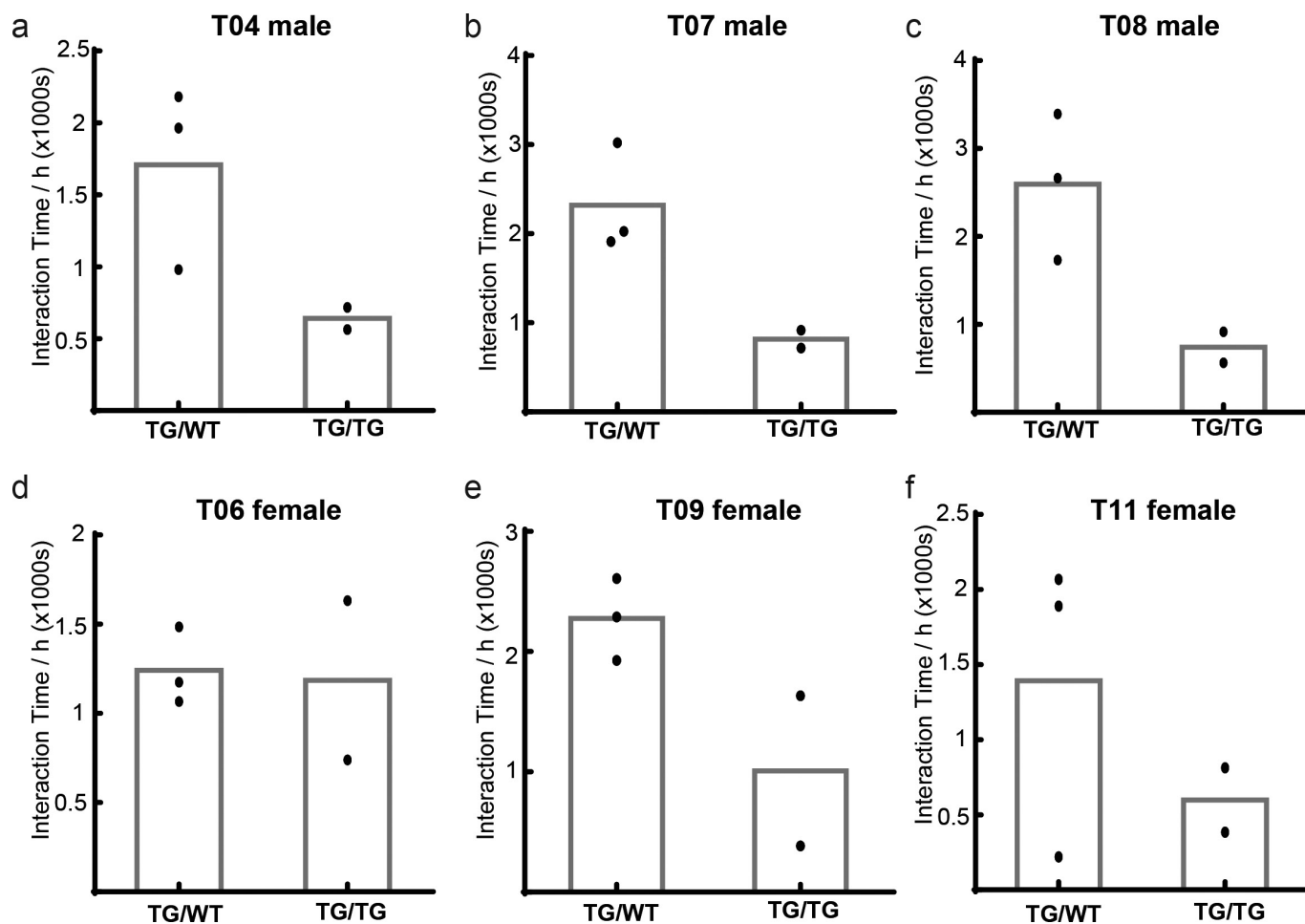


**Extended Data Figure 5 | Anxiety responses in *MECP2* TG monkeys.** **a**, Schematic illustration of the protocol of TAD test. **b**, Boxplots of the total numbers of grunt sounds made by WT and TG monkeys during the gaze period ('step 4' in the TAD test) at 36 months of age. **c**, Total grunts responses of wild-type and transgenic monkeys during TAD tests. **d**, Total vocal responses of wild-type and transgenic monkeys during TAD tests.

\* $P < 0.05$  (Student's  $t$ -test). Ends of whiskers represent the minimum and maximum of data points. The line within box represents the median (odd numbers of data points) or second quartile (even number of data points). The bottom and top edge of the box represents the first and the third quartile, respectively.

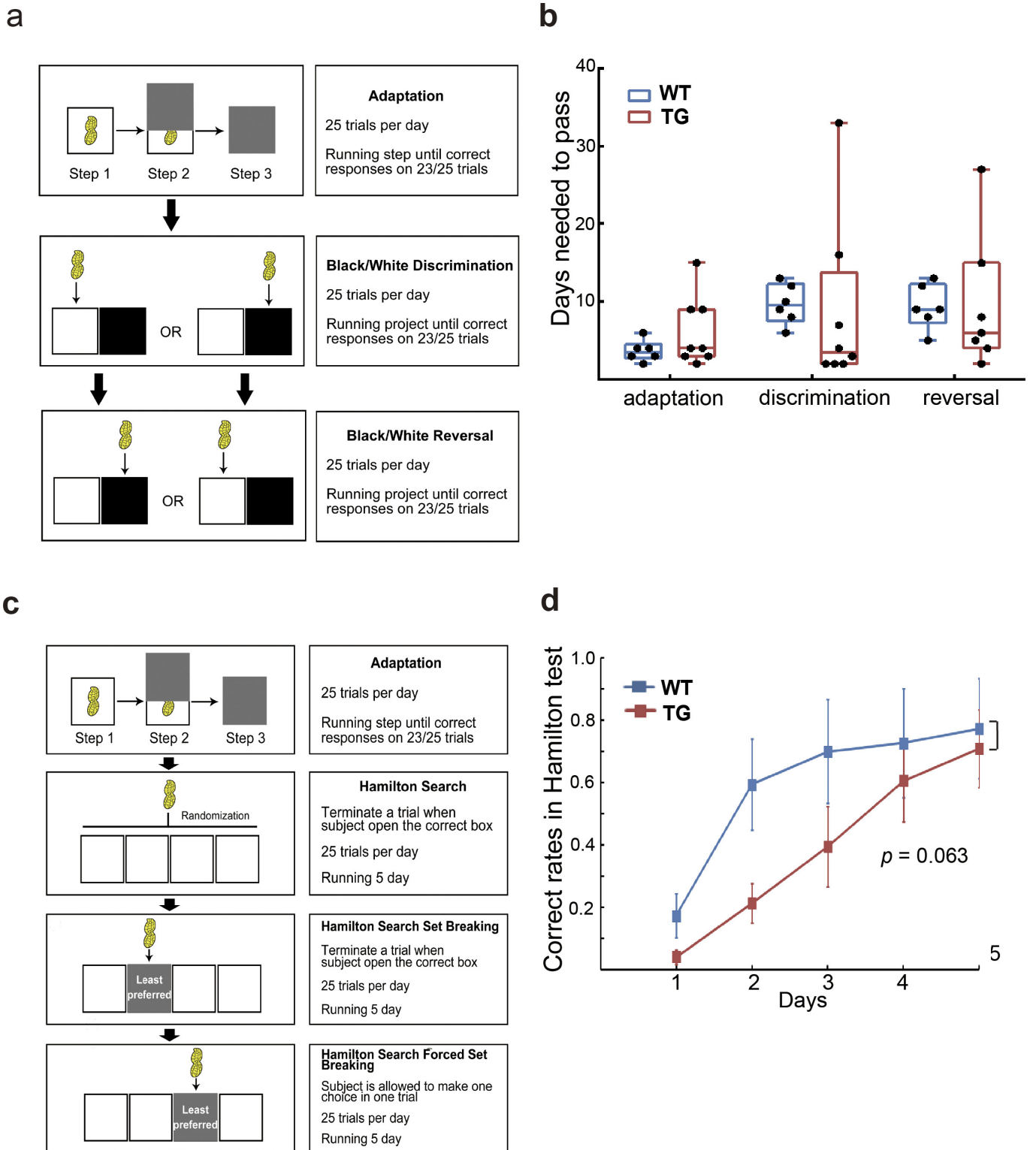


**Extended Data Figure 6 | Spectrograms of typical sounds produced by the monkey in the TAD test. a–c,** The power at different frequencies (ordinate) is colour-coded (red denotes higher power). Three typical sounds, grunt (a), coo (b) and scream (c), are shown.



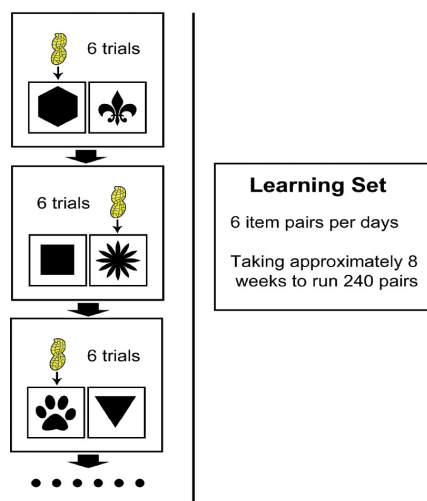
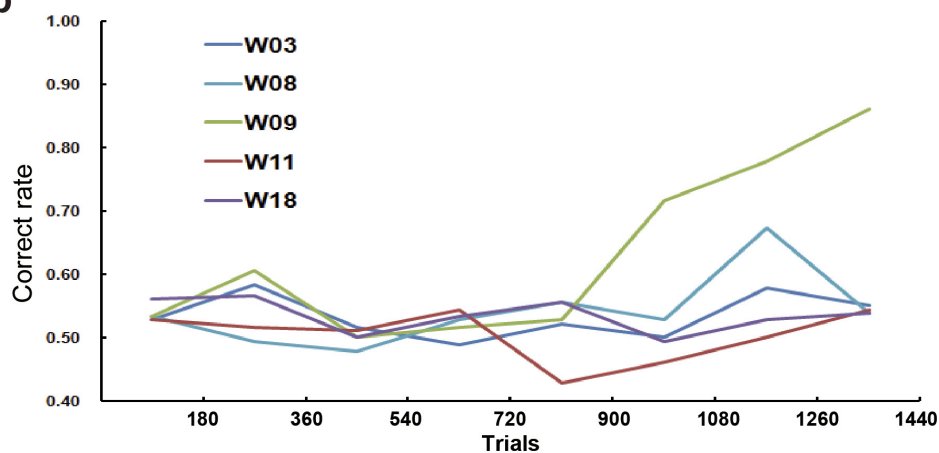
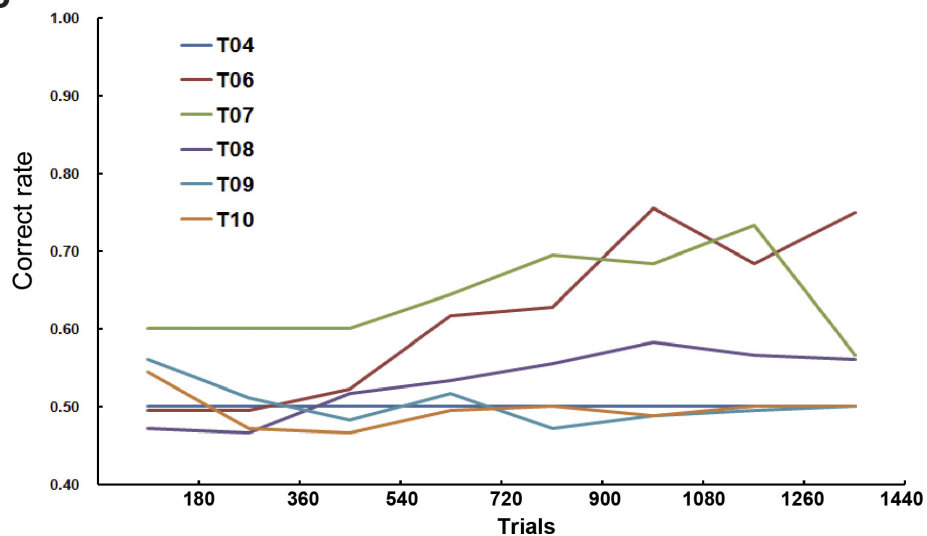
**Extended Data Figure 7 | Social interaction between monkeys from the same group (familiar pairing).** a–f, The average total time spent in sitting together during pairing in an isolated observation cage for each TG monkey (T04, T07, T08, T06, T09 and T11) with either a WT or a

TG monkey was presented individually for six TG monkeys tested. T04 (a), T07 (b), T08 (c), T06 (d), T09 (e) and T11 (f). (See Supplementary Table 4c for grouping.) Each observation lasted 60 min daily for 5 days.



data points). The bottom and top edge of the box represents the first and the third quartile, respectively. **c**, Hamilton search test. **d**, Learning curves for the Hamilton forced set-breaking test after passing the black/white test (for six WT and seven TG monkeys). The difference between the two groups was at a significance level of  $P = 0.06$  (Mann-Whitney  $U$  test). Error bars denote s.e.m.



**a****b****c**

**Extended Data Figure 9 | Performance of WT and TG monkeys in learning set of WGTA test. a,** Learning set test. Correct rate of monkeys in the reward-shape association learning test plotted individually against trials, with data points represents average correct rates over 180 trials. **b,** WT monkeys. **c,** TG monkeys.

Extended Data Table 1 | Summary of *MECP2* transgenic monkey cohorts

a.

Animal code	Date of birth (y/m/d)	Gender	Transgenes	Copy numbers identified by AccuCopy		Note
				hMECP2	mCherry	
T01	2011/2/28	F	<i>HA-hMeCP2-2a-GFP+mCherry</i>	3.4	N/A	Stillbirth
T02	2011/5/30	F	<i>HA-hMeCP2-2a-GFP+mCherry</i>	1.0	N/A	Stillbirth
T03	2011/5/30	F	<i>HA-hMeCP2-2a-GFP+mCherry</i>	N/A	N/A	Stillbirth
T04	2011/4/9	M	<i>HA-hMeCP2-2a-GFP+mCherry</i>	1.0	1.9	Live
T05	2011/3/18	F	<i>HA-hMeCP2-2a-GFP+mCherry</i>	1.2	3.8	Live, deceased at 2013/3/12
T06	2011/4/9	F	<i>HA-hMeCP2-2a-GFP+mCherry</i>	7.3	6	Live
T07	2011/4/10	M	<i>HA-hMeCP2-2a-GFP+mCherry</i>	5.7	4.3	Live, deceased at 2015/2/4
T08	2011/4/12	M	<i>HA-hMeCP2-2a-GFP+mCherry</i>	2.9	2.3	Live
T09	2011/4/16	F	<i>HA-hMeCP2-2a-GFP+mCherry</i>	2.4	3	Live, deceased at 2014/11/8
T10	2011/4/16	F	<i>HA-hMeCP2-2a-GFP+mCherry</i>	1.1	2	Live
T11	2011/6/18	F	<i>HA-hMeCP2-2a-GFP+mCherry</i>	1.9	1.3	Live
T12	2011/4/9	F	<i>mCherry</i>	0	N/A	Stillbirth

b.

Animal code	Date of birth (y/m/d)	Gender	Transgenes	Copy numbers identified by AccuCopy		Note
				hMECP2	mCherry	
TF1-1	2014/11/13	M	<i>HA-hMeCP2-2a-GFP+mCherry</i>	2.5	0.6	Deceased 3 days after birth
TF1-2	2014/11/18	M	<i>HA-hMeCP2-2a-GFP+mCherry</i>	2.5	0.5	Live
TF1-3	2014/12/7	F	<i>HA-hMeCP2-2a-GFP</i>	5.5	0	Live
TF1-4	2014/11/27	M	<i>HA-hMeCP2-2a-GFP+mCherry</i>	3.3	0.9	Live
TF1-5	2014/11/27	F	<i>HA-hMeCP2-2a-GFP+mCherry</i>	6.0	1.0	Live

a, Summary of *MECP2* F<sub>0</sub> transgenic monkeys. b, Summary of *MECP2* F<sub>1</sub> transgenic monkeys.

# The temporal scaling of *Caenorhabditis elegans* ageing

Nicholas Stroustrup<sup>1</sup>, Winston E. Anthony<sup>1</sup>, Zachary M. Nash<sup>1†</sup>, Vivek Gowda<sup>1</sup>, Adam Gomez<sup>1‡</sup>, Isaac F. López-Moyado<sup>1‡</sup>, Javier Apfeld<sup>1‡§</sup> & Walter Fontana<sup>1§</sup>

The process of ageing makes death increasingly likely, involving a random aspect that produces a wide distribution of lifespan even in homogeneous populations<sup>1,2</sup>. The study of this stochastic behaviour may link molecular mechanisms to the ageing process that determines lifespan. Here, by collecting high-precision mortality statistics from large populations, we observe that interventions as diverse as changes in diet, temperature, exposure to oxidative stress, and disruption of genes including the heat shock factor *hsf-1*, the hypoxia-inducible factor *hif-1*, and the insulin/IGF-1 pathway components *daf-2*, *age-1*, and *daf-16* all alter lifespan distributions by an apparent stretching or shrinking of time. To produce such temporal scaling, each intervention must alter to the same extent throughout adult life all physiological determinants of the risk of death. Organismic ageing in *Caenorhabditis elegans* therefore appears to involve aspects of physiology that respond in concert to a diverse set of interventions. In this way, temporal scaling identifies a novel state variable,  $r(t)$ , that governs the risk of death and whose average decay dynamics involves a single effective rate constant of ageing,  $k_r$ . Interventions that produce temporal scaling influence lifespan exclusively by altering  $k_r$ . Such interventions, when applied transiently even in early adulthood, temporarily alter  $k_r$  with an attendant transient increase or decrease in the rate of change in  $r$  and a permanent effect on remaining lifespan. The existence of an organismal ageing dynamics that is invariant across genetic and environmental contexts provides the basis for a new, quantitative framework for evaluating the manner and extent to which specific molecular processes contribute to the aspect of ageing that determines lifespan.

Body temperature is a major determinant of lifespan in poikilotherms<sup>3–5</sup> that also influences mammalian ageing<sup>6</sup>. From 20 °C to 33 °C, the mean lifespan of *C. elegans* decreases 40-fold<sup>7</sup>. To explore the impact of temperature on the actual distribution of lifespans, we used our automated imaging technology<sup>8</sup> to collect highly resolved mortality data in multiple replicate populations placed across this temperature range (Methods). From these data we estimated the survival curve  $S(t)$ , which is the probability of being alive at time (age)  $t$ , and the hazard function  $h(t) = -d \log S(t)/dt$ , which is the instantaneous risk of death at time  $t$  (Supplementary Note 1.1 and Methods).

In many invertebrates, changes in temperature alter the rate at which the risk of death increases with time<sup>4,5,9</sup>. Our lifespan data, controlled for environmental heterogeneity (see statistical methods section in Methods), confirmed this effect. However, we further observed that changes in temperature appeared to shift  $h(t)$  by an equal and opposite amount in magnitude and time when plotted on a log–log scale, suggesting that between any two temperatures  $T_0$  and  $T_1$ ,  $\lambda h_{T_1}(t) = h_{T_0}(\lambda^{-1}t)$  independent of any particular parametric form of  $h(t)$ . This change in hazard corresponds to a simple stretching of the

survival function along the time axis by a dimensionless scale factor  $\lambda$ :  $S_{T_1}(t) = S_{T_0}(\lambda^{-1}t)$  (Supplementary Note 1.2). The sole effect of changes in body temperature on lifespan therefore appeared to be a temporal rescaling of mortality statistics.

To confirm this effect, we applied an accelerated failure time (AFT) regression model<sup>10</sup> in which lifespan distributions that only differed by temporal scaling would have identically distributed residuals (Supplementary Notes 1.3 and 1.4 and Methods). To identify any significant differences between AFT residual distributions, we applied a Kolmogorov–Smirnov test adapted to censored data (Supplementary Note 2). We identified no significant temperature-dependent deviations from temporal scaling within two thermal ranges: 19–30 °C and 30.5–33 °C (Fig. 1b–d and Extended Data Figs 1–3). Populations above 30.5 °C exhibited a more pronounced late-age deceleration (Fig. 1e, Extended Data Fig. 3 and Supplementary Note 1.4), consistent with an increased heterogeneity<sup>11</sup> (Supplementary Note 3). Yet, even at high temperatures, the observed hazard function appears to be dominated more by ageing (for example, a progressive increase in the hazard) than by chance events that would produce a constant hazard (that is, non-ageing).

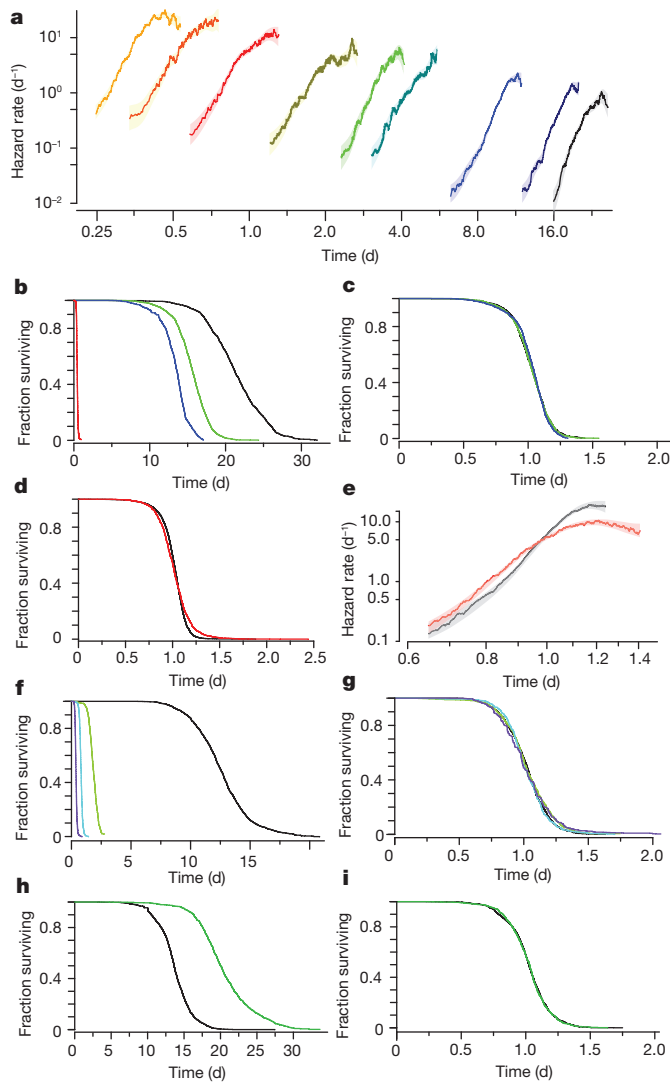
We then asked whether other interventions could produce a temporal scaling. Since oxidative damage has been linked to ageing across taxa<sup>12,13</sup>, we quantified the effect of the oxidant *tert*-butyl hydroperoxide (tBuOOH) and found that it quantitatively rescales lifespan distributions in a dose-dependent manner up to 3 mM (Kolmogorov–Smirnov  $P > 0.02$ ) with significant deviations observed only at 6 mM (Kolmogorov–Smirnov  $P = 9 \times 10^{-4}$ ; Fig. 1f–g and Extended Data Fig. 4).

To further explore the range of interventions that might yield temporal scaling, we considered three members of the insulin/IGF-1 pathway<sup>5,9</sup>: *daf-16*, a transcription factor required for lifespan extension by multiple signals<sup>14</sup>, *age-1*, a regulatory kinase upstream of *daf-16*, and *daf-2*, the insulin/IGF receptor, all of which influence both lifespan and thermal stress resistance<sup>7</sup>. Each mutant population exhibited a lifespan distribution rescaled from the wild-type distribution, both at 20 °C (Kolmogorov–Smirnov  $P > 0.015$ ; Fig. 2a–e) and at 33 °C (Kolmogorov–Smirnov  $P > 0.017$ ; Extended Data Fig. 4). The insulin/IGF receptor *daf-2* influences the activity of the heat shock factor *hsf-1* (ref. 15), and disruption of *hsf-1* also shortens lifespan by temporal rescaling (Kolmogorov–Smirnov  $P > 0.2$ ; Fig. 2c, f). Elimination of the hypoxia-inducible transcription factor *hif-1*, known to influence lifespan through *daf-16*-dependent mechanisms<sup>16</sup>, behaved likewise (Kolmogorov–Smirnov  $P > 0.2$ ; Extended Data Fig. 4).

Since changes in nutrition alter lifespan across taxa<sup>17</sup>, we considered two modifications of *C. elegans* diet: ultraviolet inactivation of the bacterial food source<sup>18</sup> and disruption of feeding behaviour by the *eat-2(ad1116)* mutation<sup>19</sup>. Ultraviolet inactivation of bacteria extended

<sup>1</sup>Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>†</sup>Present addresses: Department of Microbiology and Immunology, School of Medicine, University of North Carolina at Chapel Hill, North Carolina 27599, USA (Z.M.N.); Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, California 90095, USA (A.G.); Division of Signaling and Gene Expression, La Jolla Institute for Allergy and Immunology, La Jolla, California 92037, USA (I.F.L.-M.); Department of Biology, Northeastern University, Boston, Massachusetts 02115, USA (J.A.).

§These authors jointly supervised this work.



**Figure 1 | Environmental determinants rescale *C. elegans* lifespan distributions.** **a**, Populations grown at 20°C were transferred on their second day of adulthood to a final temperature of (right to left) 20.1°C (black), 23.7, 25.2, 29.1, 30, 30.9, 31.3, 32.5, and 32.6 (yellow). Individual lifespans were collected<sup>7</sup> and used to estimate the hazard function of each population using numerical differentiation of the Kaplan–Meier survival estimator (solid lines). The shaded areas represent the 95% confidence bands of the true hazard (Statistical methods). **d**, days. **b**, The lifespan of individuals living at 20, 25, 27, and 33°C. **c**, The data in **b** were fitted with an AFT model  $\log(y_i) = \beta x_i + \epsilon_i$  to remove differences in timescale (Methods and Supplementary Note 1.3). The AFT residuals  $\exp(\epsilon_i)$  corresponding to populations at 20, 25, and 27°C are plotted using the Kaplan–Meier survival estimator. **d**, The AFT residuals corresponding to populations held at 25 (black) and 33°C (red) are plotted using the Kaplan–Meier survival estimator. **e**, Hazard functions were estimated from the 25 and 33°C AFT residuals. **f**, The survival curves of populations exposed to 0 (black), 1.5 (blue), 3 (green), and 6 mM (purple) tBuOOH. **g**, The AFT residuals for the data of **f**. **h**, The survival curves of animals cultured on live *E. coli* (black) and ultraviolet-inactivated *E. coli* (green). **i**, The AFT residuals for the data of **h**.

lifespan via temporal scaling (Kolmogorov–Smirnov  $P > 0.2$ ; Fig. 1h, i). In contrast, *eat-2(ad1116)* populations exhibited a significant deviation from temporal scaling (Kolmogorov–Smirnov  $P = 5 \times 10^{-5}$ ), with a disproportionate increase in the standard deviation of lifespan compared with the mean (Fig. 2g, j). We also noted that *eat-2(ad1116)* populations exhibited a substantially increased variation in developmental timing. While such variation does not affect lifespan statistics based on manually synchronized young adults (Methods), it is

possible that the causes of this developmental variation also underlie the increased variation of lifespan. We found that disruption of the mitochondrial complex I in *nuo-6(qm200)* populations produced analogous effects on developmental timing with a deviation from temporal scaling of lifespan similar to *eat-2(ad1116)* (Kolmogorov–Smirnov  $P > 3 \times 10^{-18}$ ; Fig. 2h, k). Yet, populations with either allele exhibited temporally rescaled lifespan distributions in response to temperature changes (Kolmogorov–Smirnov  $P > 0.2$ ; Fig. 2i, l and Extended Data Fig. 4). We conclude that while *eat-2(ad1116)*, *nuo-6(qm200)*, and shifts in temperatures from below to above 30°C alter lifespan distributions outside the temporal scaling model, these interventions do not eliminate the ability of *C. elegans* to respond to subsequent interventions with temporal scaling. Temporal scaling thus appears to be a pervasive response to interventions of diverse modality and intensity.

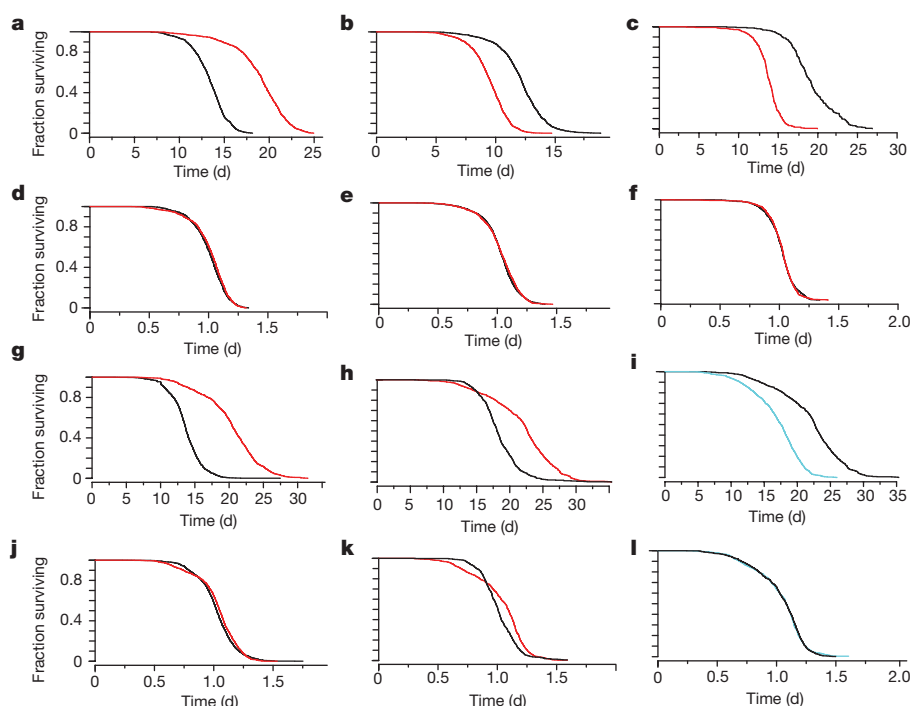
Temporal scaling would arise if all physiological determinants of the risk of death in *C. elegans* acted as if they were jointly governed by a single stochastic process whose rate constant alone was altered by interventions (Supplementary Note 4). If the risk of death was determined in this way, we reasoned that transient interventions early in adulthood would produce a persistent temporal shift, not a scaling, of mortality statistics (Supplementary Note 4.3). To test this, we focused on temperature, which can be quantitatively, rapidly, and reversibly switched at any age between a baseline temperature  $T_0$  and a transient temperature  $T_1$  (Fig. 3a). We confirmed that transient exposure to higher temperatures produced a permanent shortening of lifespan<sup>5</sup> (Fig. 3b). We found that this shortening consisted of a temporal shift of the lifespan distribution (Fig. 3c, d)  $S_{T_1}(t) = S_{T_0}(t - \Delta_\tau)$  that matches the magnitude of shift  $\Delta_\tau$  predicted if time were rescaled only for the period  $\tau$  that animals were held at the transient temperature:  $\Delta_\tau = \tau(1 - \lambda^{-1})$ , with  $\lambda$  the scale factor relating populations always held at  $T_1$  to populations always held at  $T_0$  (Fig. 3e, f, Supplementary Note 4.3, Supplementary Table 2 and Extended Data Fig. 5). In a complementary experiment, we found that exposure to high temperature for different periods  $\tau$  also gave shifts with the predicted magnitude (Extended Data Fig. 5). It appears, therefore, that the temporal scaling observed in Fig. 1a and the temporal shifting of Fig. 3 are compatible with a single model in which interventions alter the effective rate constant of a stochastic process governing those aspects of *C. elegans* physiology that determine risk of death. This process is evidently ongoing even very early in adulthood and is governed by the same rate constant as in late adulthood.

To clarify how molecular pathways contribute to temporal scaling, we quantified the magnitude of scaling produced by different intensities of intervention: that is, the scaling function. In the case of temperature, we applied an Arrhenius analysis<sup>20,21</sup> to interpret the change of  $\lambda$  (which in our framework rescales the rate constant of ageing) across the range 20–35°C (Fig. 4a). We identified three distinct thermal regimes: I, 20–29.4°C; II, 29.4–32.1°C; III, 32.1–35°C (Fig. 4b, Methods and Extended Data Figs 6 and 7) with regime I being further subdivided into Ia and Ib by a reproducible transition point at 24.4°C.

Each scaling regime appears to correspond to a distinct molecular mechanism and barrier process dominating the timescale of ageing (Supplementary Table 1). Sharp decreases in lifespan have been observed to occur around 30°C in *Drosophila melanogaster*<sup>21</sup>, hinting at a more general phenomenon in poikilotherms. Notably, this transition coincided with a deviation from temporal scaling of lifespan distributions (Fig. 1e and Extended Data Fig. 3). Intriguingly, the scaling across the breakpoint between regimes Ia and Ib suggested that temporal scaling need not be disrupted by a change in the molecular mechanisms dominating the timescale of ageing.

Quantifying the effects of temperature on mutant strains, we found that the elimination of DAF-16 shortened lifespan by a rescaling of 28% in regime Ia and 25% in Ib (Fig. 4c, d). The *daf-16(mu86)* population exhibited the same slope in scaling function as wild type in Ia, and differed only by about 5% across regime Ib, suggesting that the mechanisms mediating the temperature dependence of lifespan





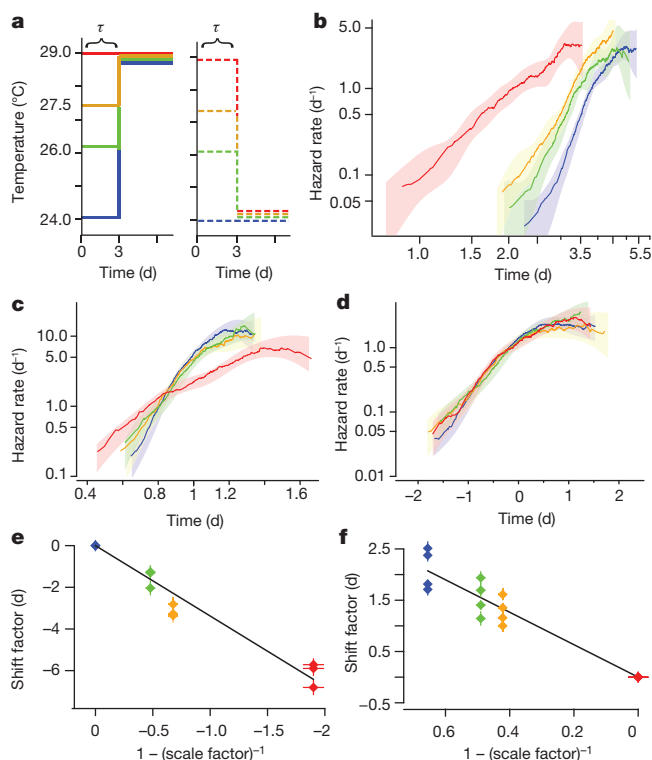
**Figure 2 | Genetic determinants rescale *C. elegans* lifespan distributions.** **a–c**, Survival curves are shown for *daf-2(e1368)* (red) and wild type (black) at 25 °C (**a**), *daf-16(mu86)* (red) and wild type (black) at 25 °C (**b**), and *hsf-1(sy441)* (red) and wild type (black) at 20 °C (**c**). **d–f**, The AFT residuals corresponding to the data in **a–c** respectively.

Survival curves are shown for *eat-2(ad1116)* (red) and wild type (black) at 20 °C (**g**), *nuo-6(qm200)* (red) and wild type (black) at 25 °C (**h**), and *nuo-6(qm200)* populations held at 20 °C and 25 °C (**i**). **j–l**, The AFT residuals corresponding to the data in **g–i** respectively.

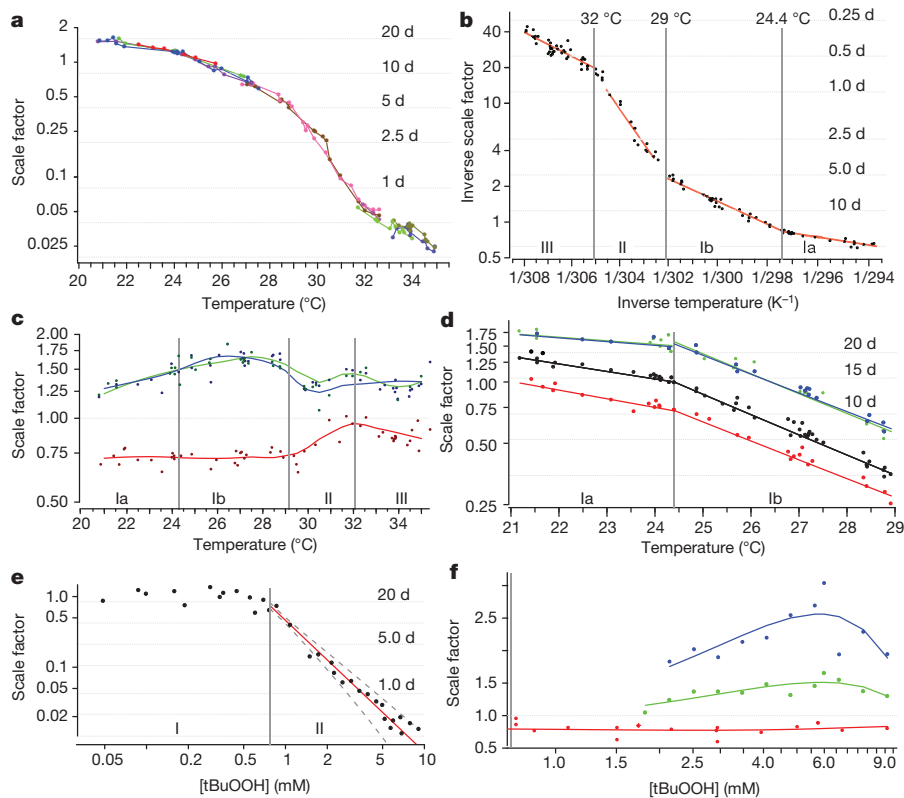
in regime I were not altered by elimination of DAF-16. In contrast, the hypomorphic alleles *daf-2(e1368)* and *age-1(hx546)* exhibit clear temperature-dependent effects across regime I (Fig. 4c, d). Both genes influence lifespan at 20 °C and 35 °C primarily by suppressing *daf-16* activity<sup>22</sup>, which itself appears independent of temperature. Thus, *daf-2(e1368)* and *age-1(hx546)* alleles appear to be neomorphic

in respect of the temperature dependence of their regulation of DAF-16.

We found that tBuOOH decreased lifespan at concentrations above 750  $\mu$ M, with  $\lambda$  decreasing as a power law (Fig. 4e and Methods). This suggests an overall mass-action kinetics for the chain of events linking the direct targets of tBuOOH to the rescaling of the lifespan



**Figure 3 | Transient interventions during early adulthood shift the lifespan distribution.** **a**, A schematic: populations were placed at 24 °C (blue), 26 °C (green), 27.5 °C (orange), and 29 °C (red). After  $\tau = 3.2$  days, sub-populations were transferred to either 24 °C or 29 °C for the remainder of their lives. **b**, The hazard rate was estimated using the remaining lifespan of populations transferred to the final temperature of 29 °C. **c**, To test for temporal scaling between the populations shown in **b**, death times were fitted with the regression model  $\log(y_i) = \beta x_i + \epsilon_i$ , in which  $\exp(\beta_i)$  is the best estimate for the scale factor  $\lambda$ . The residuals  $\exp(\epsilon_i)$  are plotted as hazard functions in the colour scheme of **a**. **d**, To test for temporal shifts between the populations shown in **b**, death times were fitted with the regression model  $y_i = \beta x_i + \epsilon_i$ , in which  $\beta_i$  is the best estimate for the shift term  $\Delta_\tau$ . The residuals  $\epsilon_i$  are plotted as hazard functions in the colour scheme of **a**. **e**, The shift term  $\Delta_\tau$  for populations transferred from each high temperature to 24 °C was plotted against  $1 - \lambda^{-1}$ , where  $\lambda$  is the scale factor relating populations always held at the corresponding high temperature to those always held at 24 °C. The prediction  $\Delta_\tau = \tau(1 - \lambda^{-1})$  suggests that these points should fall along a line with a slope equal to  $\tau$  in **a**. A linear regression on these points model estimates  $\tau = 3.38 \pm 0.17$ . **f**, As in **e**, but for populations transferred from lower initial temperatures to the final higher temperature of 29 °C, producing the estimate  $\tau = 3.16 \pm 0.14$ .



**Figure 4 | Scaling functions.** **a**, The magnitude of temporal scaling was estimated for wild-type populations held at fractional degree intervals across the range 20–35 °C. The scale factor  $\lambda$  of each population was estimated relative to a reference population at 25 °C. Grey lines mark the average lifespan of the reference population scaled by  $\lambda$ . Each replicate is shown as a separate colour, with each point corresponding to an aggregate population consisting of on average 130 individuals at the outset. **b**, The scale factor  $\lambda$  was determined for populations across the temperature range of **a**. The data points were fitted with a segmented Arrhenius model  $\lambda(T)^{-1} = p_0 \exp(-p_1/RT)$  (red). **c**, The magnitude of scaling produced by *daf-16(mu86)* (red), *daf-2(e1368)* (green), and *age-1(hx546)* (blue) alleles

relative to wild type was estimated at each temperature considered (points). Solid curves represent trends across temperature as fitted by a Loess regression. **d**, The combined magnitude of scaling produced by each allele and change of temperature was estimated relative to a single wild-type population 24 °C; colours as in **c**. Regimes II and III are shown in Extended Data Fig. 7. **e**, Wild-type populations at 20 °C were exposed to a series of tBuOOH concentrations ranging from 0 to 10 mM. For each population,  $\lambda$  was calculated relative to an unexposed population (0 mM). Data for concentrations above 0.75 mM were fitted by the model  $\lambda([tBuOOH]) = p_2([tBuOOH])^{p_3}$  (red), yielding  $p_2 = 0.47 \pm 0.02$  and  $p_3 = -1.86 \pm 0.15$ . **f**, As in **c**, but for the tBuOOH dosage series.

distribution. The distinct scaling functions of tBuOOH (power law) and temperature (multiple Arrhenius regimes) further suggest distinct molecular targets and mechanisms through which each type of intervention rescales the lifespan distribution.

As with temperature, the elimination of DAF-16 in the presence of tBuOOH reduced lifespan by a constant amount (Fig. 4f),  $19.5 \pm 8.8\%$ , across all concentrations tested. Taken together with our temperature data in Fig. 4c, these results suggest that DAF-16 acts antagonistically but in parallel to the mechanisms through which tBuOOH and temperature shorten lifespan. DAF-16, tBuOOH, and temperature appear to affect ageing through their influence on risk determinants downstream of all three. For example, DAF-16 might attenuate or mitigate certain types of error or damage regardless of how the errors are created. The magnitude of temporal scaling produced both by *daf-2(e1368)* and by *age-1(hx546)* alleles varied across tBuOOH concentrations (Fig. 4g), which seems yet another aspect of a quantitative stress-dependent regulation of DAF-16 present in these strains but absent in wild type.

Disruption of *daf-2*, *daf-16*, *hif-1* or *hsf-1* produces distinct metabolic, cell-biological, and behavioural effects<sup>15,23</sup>, as do changes in diet<sup>24</sup>, temperature<sup>25</sup>, and exposure to tBuOOH<sup>26</sup>. Yet, temporal scaling arises independently of the molecular targets specific to each intervention and requires that all risk determinants be affected to the same extent. This suggests that ageing in *C. elegans* can be described in terms of a whole-organism state variable  $r$  that completely determines all-cause mortality (Extended Data Fig. 9). State variables familiar from

other contexts include temperature, pressure, and entropy, all of which describe the behaviour of a system resulting from the collective action of its many constituent elements without reference to their nature. In the same way, the change of the state over time,  $r(t)$ , describes the ageing process of *C. elegans* in terms of a collective action of all physiological determinants of risk. Where multiple risk determinants independently influence lifespan, temporal scaling requires that interventions simultaneously rescale, to an identical extent throughout life, the risk functions associated with each determinant (Supplementary Note 5.1). In models including dependencies among risk determinants, temporal scaling can emerge even when interventions act differentially across risk determinants (Supplementary Notes 5.2 and 5.3): dependencies can propagate the influence of interventions from one to all risk determinants, in effect producing a system-wide property that we call  $r(t)$ .

The temporal scaling of lifespan distributions constrains the dynamics of the state variable  $r(t)$ : the single stochastic process determining *C. elegans* lifespan must be invariant to timescale transformations and follow an average dynamics governed by an effective rate constant:  $dr/dt = -k_r F(r)$ , where  $F(r)$  is an unknown function of  $r$  that does not depend on  $k_r$ . In this formulation, temporal scaling arises when interventions change  $k_r$  into  $k_r/\lambda$ . These dynamics place constraints on any stochastic process proposed to describe organismal ageing, as its parameters must change in a coordinated fashion. For example, if  $r(t)$  were described by a biased random walk<sup>27</sup>, the drift coefficient and the square of the diffusion coefficient must remain in a fixed proportion under intervention (Supplementary Note 6).

The idea that ageing is driven by changes in an organismal physiological state has been variously framed in terms of notions such as organization, vitality, organ reserve or resilience<sup>3,28,29</sup>. The temporal scaling across interventions justifies this notion, allowing an initial formalization. We note that any aspects of *C. elegans* physiology that change over time but do not influence lifespan, influencing 'quality' rather than 'quantity' of life, need not change in concert with  $r(t)$ .

We know neither the physiological basis of the state  $r(t)$  nor the specific dynamics by which it changes with age. Yet, we can expect a broad set of lifespan determinants to affect only  $k_r$ , including minimally all determinants that influence lifespan exclusively through DAF-16 (refs 14 and 30), HSF-1 or HIF-1, or through the mechanisms that mediate the effects of temperature and tBuOOH on lifespan. If most ageing mechanisms currently studied influence only  $k_r$ , then future studies directed at clarifying the physiological origins of  $r$  and its dynamics should identify novel ageing mechanisms  $F(r)$ .

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 27 July; accepted 18 December 2015.**

**Published online 27 January 2016.**

- Pincus, Z., Smith-Vikos, T. & Slack, F. J. MicroRNA predictors of longevity in *Caenorhabditis elegans*. *PLoS Genet.* **7**, e1002306 (2011).
- Herndon, L. A. *et al.* Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*. *Nature* **419**, 808–814 (2002).
- Shaw, R. F. & Bercaw, B. L. Temperature and life-span in poikilothermous animals. *Nature* **196**, 454–457 (1962).
- Mair, W., Goymier, P., Pletcher, S. D. & Partridge, L. Demography of dietary restriction and death in *Drosophila*. *Science* **301**, 1731–1733 (2003).
- Wu, D., Rea, S. L., Cypser, J. R. & Johnson, T. E. Mortality shifts in *Caenorhabditis elegans*: remembrance of conditions past. *Ageing Cell* **8**, 666–675 (2009).
- Conti, B. *et al.* Transgenic mice with a reduced core body temperature have an increased life span. *Science* **314**, 825–828 (2006).
- Lithgow, G. J., White, T. M., Melov, S. & Johnson, T. E. Thermotolerance and extended life-span conferred by single-gene mutations and induced by thermal stress. *Proc. Natl Acad. Sci. USA* **92**, 7540–7544 (1995).
- Stroustrup, N. *et al.* The *Caenorhabditis elegans* lifespan machine. *Nature Methods* **10**, 665–670 (2013).
- Johnson, T. E., Wu, D., Tedesco, P., Dames, S. & Vaupel, J. W. Age-specific demographic profiles of longevity mutants in *Caenorhabditis elegans* show segmental effects. *J. Gerontol. A* **56**, B331–B339 (2001).
- Swindell, W. R. Accelerated failure time models provide a useful statistical framework for ageing research. *Exp. Gerontol.* **44**, 190–200 (2009).
- Vaupel, J. W., Manton, K. G. & Stallard, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454 (1979).
- Martin, G. M., Austad, S. N. & Johnson, T. E. Genetic analysis of ageing: role of oxidative damage and environmental stresses. *Nature Genet.* **13**, 25–34 (1996).
- Tullet, J. M. *et al.* Direct inhibition of the longevity-promoting factor SKN-1 by insulin-like signaling in *C. elegans*. *Cell* **132**, 1025–1038 (2008).
- Landis, J. N. & Murphy, C. T. Integration of diverse inputs in the regulation of *Caenorhabditis elegans* DAF-16/FOXO. *Dev. Dyn.* **239**, 1405–1412 (2010).
- Hsu, A. L., Murphy, C. T. & Kenyon, C. Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science* **300**, 1142–1145 (2003).
- Leiser, S. F., Begun, A. & Kaeblerlein, M. HIF-1 modulates longevity and healthspan in a temperature-dependent manner. *Ageing Cell* **10**, 318–326 (2011).
- Kirkwood, T. B. & Shanley, D. P. Food restriction, evolution and ageing. *Mech. Ageing Dev.* **126**, 1011–1016 (2005).
- Garigan, D. *et al.* Genetic analysis of tissue aging in *Caenorhabditis elegans*: a role for heat-shock factor and bacterial proliferation. *Genetics* **161**, 1101–1112 (2002).
- Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95**, 13091–13096 (1998).
- Suda, H., Sato, K. & Yanase, S. Timing mechanism and effective activation energy concerned with aging and lifespan in the long-lived and thermosensory mutants of *Caenorhabditis elegans*. *Mech. Ageing Dev.* **133**, 600–610 (2012).
- Altan, H., Miquel, J., Helmle, L. C. & Dolcas, C. B. Thermodynamics of aging in *Drosophila melanogaster*. *Mech. Ageing Dev.* **5**, 371–387 (1976).
- Libina, N., Berman, J. R. & Kenyon, C. Tissue-specific activities of *C. elegans* DAF-16 in the regulation of lifespan. *Cell* **115**, 489–502 (2003).
- Murphy, C. T. *et al.* Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424**, 277–283 (2003).
- Walker, G., Houthoofd, K., Vanfleteren, J. R. & Gems, D. Dietary restriction in *C. elegans*: from rate-of-living effects to nutrient sensing pathways. *Mech. Ageing Dev.* **126**, 929–937 (2005).
- McColl, G. *et al.* Insulin-like signaling determines survival during stress via posttranscriptional mechanisms in *C. elegans*. *Cell Metab.* **12**, 260–272 (2010).
- Oliveira, R. P. *et al.* Condition-adapted stress and longevity gene regulation by *Caenorhabditis elegans* SKN-1/Nrf. *Ageing Cell* **8**, 524–541 (2009).
- Aalen, O. O. & Gjessing, H. K. Understanding the shape of the hazard rate: a process point of view. *Stat. Sci.* **16**, 1–14 (2001).
- Medawar, P. B. *An Unsolved Problem of Biology: An Inaugural Lecture Delivered at University College, London* (H. K. Lewis, 1951).
- Gladyshev, V. N. The origin of aging: imperfectness-driven non-random damage defines the aging process and control of lifespan. *Trends Genet.* **29**, 506–512 (2013).
- Samuelson, A. V., Carr, C. E. & Ruvkun, G. Gene activities that mediate increased life span of *C. elegans* insulin-like signaling mutants. *Genes Dev.* **21**, 2976–2994 (2007).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. Alcedo for nematode strains, X. Manière for providing the NEC937 *Escherichia coli* strain, B. Ward for reading our manuscript, and D. Marks, C. Romero, T. Kolokotronis, D. Yamins, P. F. Stadler, E. Smith, and all members of the Fontana laboratory for discussions and encouragement throughout this project. Some strains were provided by the *Caenorhabditis* Genetics Center, which is funded by US National Institutes of Health (NIH) Office of Research Infrastructure Programs (P40 OD010440). This work was funded by the NIH through grant R01 AG034994 and by a Glenn Award from the Glenn Foundation for Medical Research.

**Author Contributions** N.S. conceived and analysed the experiments. N.S., J.A., W.E.A., V.G., A.G., Z.M.N., and I.F.L.-M. developed experimental protocols and performed experiments. N.S. and W.F. interpreted data and performed model calculations. N.S. and W.F. wrote the manuscript with input from J.A.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.S. ([nstroustrup@post.harvard.edu](mailto:nstroustrup@post.harvard.edu)) or W.F. ([walter@hms.harvard.edu](mailto:walter@hms.harvard.edu)).



## METHODS

**Experimental methods.** The following nematode strains were used: QZ0 (wild type (Bristol)), TJ1052 (*age-1(hx546)II*), and QZ120 (*daf-2(e1368)*), QZ60 (*daf-16(mu86)*), QZ121 (*hsf-1(sy441)*), QZ414 (*eat-2(ad1116)*), ZG31 (*hif-1(ia4)*), and MQ1333 (*nuo-6(qm200)*).

Hermaphrodites were cultured under standard conditions<sup>8,31</sup>, at either 20 or 25 °C as noted, on plates containing 100 µg ml<sup>-1</sup> ampicillin and seeded at an absorbance at 600 nm of 20 with the *E. coli* strain NEC937 B (OP50 ΔuvrA; KanR)<sup>32</sup>. Before seeding, all bacteria were irradiated in batch culture with 4 J m<sup>-2</sup> of 254-nm light in a UV Stratalinker (Stratagene). Age-synchronous cohorts were prepared by hypochlorite treatment<sup>33</sup>. In strains exhibiting an increased variation in developmental time, in particular *eat-2(ad1116)* and *nuo-6(qm200)*, developmentally synchronized L4 larvae were manually selected at the final stage of vulval maturation. To control for any effects of temperature on *C. elegans* development, all populations were held at 20 °C until their second day of adulthood, at which time they were exposed to the interventions described (unless otherwise stated). Individuals were randomly assigned groups by obtaining aliquots from populations suspended in M9 buffer. Automated lifespan experiments were run, blinded, and validated according to published methods<sup>8</sup>.

All populations were transferred on their second day of adulthood onto plates containing 22.5 µg ml<sup>-1</sup> nystatin (Sigma N3503) to prevent fungal growth and 27.5 µg ml<sup>-1</sup> 5-fluoro-2-deoxyuridine (FUDR, Sigma) to eliminate live progeny. Where live bacteria were used (Fig. 1h, i), 10 µg ml<sup>-1</sup> FUDR sufficed to eliminate live progeny.

Standard NGM agar plates were poured and dried according to published methods<sup>8</sup>. For tBuOOH assays, the compound was added to molten agar immediately before pouring. All plates containing tBuOOH were seeded and placed in a fume hood until the bacteria was absorbed, approximately 1 h. We found no evidence of time-dependent degradation of tBuOOH in solid agar plates, nor any effect of the tBuOOH degradation byproduct tbutanol on *C. elegans* lifespan (Extended Data Fig. 4), although evaporation of tBuOOH needed to be strictly controlled.

Scanner temperatures were measured using thermocouples (ThermoWorks USB-REF) mounted on the bottom of empty Petri dishes loaded onto each scanner. Scanners were given several hours to reach their stable operating temperature, at which point measurements were taken every 10 s. Because operating scanners exhibit small, regular oscillations in temperature<sup>8</sup>, the average of multiple cycles, each lasting 20 min, was taken.

**Statistical methods.** *Population sizes.* Supplementary Table 2 shows various summary statistics, including population size, for the population in each figure panel. No statistical methods were used to predetermine sample size. Initial experiments showed that populations larger than 500 individuals provided more robust survival estimates. Statistical power was estimated retrospectively (Supplementary Notes 2.2 and 6.3). Animals lost from observation were censored according to published methods<sup>8</sup>.

*Replicates.* All replicates described are biological replicates, performed in separate weeks using separate populations. All replicates performed in the laboratory yielding informative population sizes and meeting self-consistency standards (for example consistent results across multiple scanners<sup>8</sup>) are shown.

*Single AFT regression (all figures).* Using our automated microscopy method<sup>8</sup>, a single, very large, homogeneous population must be distributed across multiple scanners. The local environment characteristic to each scanner can influence *C. elegans* lifespan. This influence can be measured and controlled using the same AFT model we use to quantify temporal scaling. In most cases, our AFT models take the form

$$\log(y_i) = \beta X_i + \epsilon_i \quad (1)$$

where  $y_i$  is the lifespan of individual  $i$ ,  $\beta$  is the parameter vector being estimated, and  $X_i$  is the categorical covariate coding for the label associated with each individual  $i$ . This label takes different values in different contexts, representing either the plate name (Figs 1, 2, 3b, d and 4b–f), scanner name (Fig. 4a) or animal genotype (Fig. 4c, f). We take the logarithm of lifespan,  $\log(y_i)$ , following the standard approach for evaluating covariates that act multiplicatively, producing fold-changes in lifespan.

Each AFT regression model has an intercept that determines the ‘reference’ lifespan in relation to which all parameter vectors  $\beta$  are scaled. The intercept is implicitly determined by the particular categorical encoding scheme used for  $X_i$ . In Figs 1, 2 and 3b–d a ‘deviation’ coding is used, placing the intercept at the grand mean lifespan of all individuals. In Fig. 4, a dummy coding is used, placing the intercept at the mean lifespan of a specific reference group whose identity is noted in each case. The coding scheme used for the categorical variable has no effect on the value of model residuals  $\epsilon_i$  and determines only the values of the AFT parameters  $\beta$ .

All AFT regression models were estimated via Buckley–James regression (included in the R<sup>34</sup> package RMS). In Figs 1a and 3b and Extended Data Fig. 5a, j, l, n, p, r, t, v ‘device-corrected death times’<sup>8</sup> were calculated to remove obscuring effects of temperature variation between scanners on the aggregate hazard rate. A dummy encoding for  $X_i$  coded for the scanner on which each individual was observed. The model intercept was then added to the residuals  $\epsilon_i$ .

In many panels, AFT residuals  $\epsilon_i$  are grouped and plotted according to some experimental condition: by temperature in Figs 1c, d, e and 2l and Extended Data Fig. 4i; by tBuOOH concentration in Fig. 1g and Extended Data Fig. 4a; by bacterial treatment in Fig. 1i; and by genotype in Fig. 2d, e, f, j, k and Extended Data Fig. 4e, g, k, m, o, q, s, u, w. We emphasize that in each case, the experimental condition used to group plots is not a covariate in the AFT model. Instead, the single AFT categorical covariate codes for an experimental unit—plate or scanner name—as detailed above. In this way, any temporal rescaling produced by variation in environmental conditions across these units will be estimated and reflected in the parameter vector  $\beta$ , and its effects minimized on the residual times  $\epsilon_i$ . These residuals are then grouped according to the experimental variable during subsequent analysis (Supplementary Note 2).

*Single AFT regression for additive models (Fig. 3d–f, h and Extended Data Fig. 5b, e, l).* To account for temporal shifts in mortality statistics, Buckley–James regression (using the R<sup>34</sup> package RMS) was used to fit the model

$$y_i = \beta X_i + \epsilon_i \quad (2)$$

where  $X_i$  is the categorical covariate the initial (Fig. 3d–f and Extended Data Fig. 5b) temperature at which individual  $i$  was placed, the duration of time spent at that temperature (Extended Data Fig. 5d–g), or the number of switches used (Extended Data Fig. 5l). A ‘deviation’ coding was used (see equation (1)) in Fig. 3d and a dummy coding was used for Fig. 3e, f.

*Significance (Figs 1 and 2).* The probability ( $P$  value) that the observed differences in lifespan between two populations are explained entirely by temporal scaling was estimated using the two-sided modified Kolmogorov–Smirnov test to identify heteroscedasticity among AFT residuals. This approach is described, with additional power analysis, in Supplementary Note 2.

*Hierarchical clustering of survival curves (Extended Data Fig. 3).* A hierarchical clustering of survival curves was computed to identify groups of curves temporally rescaled in respect to each other. All pairs of populations were compared using the modified Kolmogorov–Smirnov test as described in Supplementary Note 2. The modified Kolmogorov–Smirnov distance (Supplementary Note 2) was used as the distance metric for clustering, using the R<sup>34</sup> hierarchical clustering implementation *hclust*.

*Estimating hazard functions (Figs 1–3).* The time-dependent hazard rate was estimated through numerical differentiation of the Kaplan–Meier cumulative hazard estimate. To generate confidence bands for the true hazard rate under the assumption that it is locally smooth, death times were fitted with a piecewise-polynomial B-spline hazard model using the R<sup>34</sup> package *bshazard*<sup>35</sup>.

*Estimating the magnitude of temporal scaling across temperatures (Fig. 4a, b).* Buckley–James regression was used to fit the AFT model described in equation (1), with  $X_i$  as a categorical covariate coding the scanner name corresponding to each individual tested. By specifying a single scanner at 25 °C as the reference category of a reference coded categorical variable, the AFT parameter vector  $\beta$  becomes the best estimate of the scale factors relating lifespan on each scanner  $X_i$  to the 25 °C reference:  $S_X(t) = S_{25^\circ\text{C}}(\lambda^{-1}t)$ .

Because the average temperature was measured for each scanner  $X_i$ , the corresponding  $\lambda$  can therefore be plotted in Fig. 4a, b as a function of the temperature on the scanner that produced it.

*Identifying distinct scaling regimes (Fig. 4b and Extended Data Figs 6 and 7a, b).* Because the temperature scaling of timescale  $\lambda$  did not appear uniform, we applied a linear segmented regression model using transformed variables to identify the number scaling regimes and estimate the boundaries between them. An Arrhenius model was fitted, assuming that within each segment

$$\lambda(T)^{-1} = Ae^{-B/T} \quad (3)$$

with  $\lambda(T)$  as the scale factor at temperature  $T$ , and  $A$  and  $B$  as pre-exponential and exponential Arrhenius constants, respectively. One or more breakpoints were incorporated at temperatures  $T = \{T_1, \dots, T_{n-1}\}$  to produce the segmented model

$$\log(\lambda(T)^{-1}) = A_i - \frac{B_i}{T}, T_{i-1} < T < T_i, i = 1, \dots, n \quad (4)$$

with  $T_0$  and  $T_n$  the fixed starting and end temperatures, respectively. A linear model was tested in the same fashion:

$$\lambda(T)^{-1} = A_i + B_i/T, T_{i-1} < T < T_i, i = 1, \dots, n \quad (5)$$

Model parameters were estimated using the R<sup>34</sup> package ‘segmented’.



To generate best estimates for the  $B$  parameter within each Arrhenius scaling regime identified by segmented regression, death times within each regime were isolated and fitted using a nonlinear regression approach (R<sup>34</sup> package nls2). In all segmented and nonlinear regression models performed on  $\lambda$  values, each  $\lambda$  was weighted according to the size of the population used in the AFT regression from which the  $\lambda$  was estimated. Additional estimates of Arrhenius parameters were obtained following the approach described in equation (6).

*Estimating the effect of mutations relative to wild type, across temperatures (Fig. 4c).* To estimate the effect of mutant alleles on lifespan relative to wild type at each temperature, a separate AFT model as specified in equation (1) was run at each temperature considered. In each regression,  $X_i$  was specified as the genotype of individuals, with the categorical variable coded to use wild-type populations as the reference.

*Characterizing the temperature dependence of mutant lifespan across temperatures, relative to wild type (Fig. 4d and Extended Data Fig. 7c).* To identify differences in the temperature dependence of lifespan of mutants relative to wild type, we considered each thermal regime separately (regimes Ia and Ib in Fig. 4d and regimes II and III in Extended Data Fig. 7). We considered the model

$$\log(y_i) = \beta_g X_i + \beta_r Y_i + \beta_T T_i + \beta_{XT} X_i T_i + \epsilon_i \quad (6)$$

where  $X_i$  is a categorical variable encoding for the genotype of the individual,  $Y_i$  is the biological replicate in which the individual was observed (animals originating from the same hypochlorite treatment in the same week have the same  $Y_i$ ), and  $T_i$  is a continuous variable representing the temperature (in inverse degrees Kelvin) at which the individual was placed. The interaction term  $\beta_{XT} X_i T_i$  captures any differential effect of temperature on the mutant strain relative to wild type. The null hypothesis, therefore, is that mutant individuals exhibit no difference in their response to temperature compared with wild-type individuals, which can be rejected if a significant non-zero value of  $\beta_{XT}$  is observed, as calculated from the associated Wald Z score. The estimated model parameters are provided in Supplementary Table 3.

*Multiple AFT regression across tBuOOH replicates (Fig. 4e).* The absolute effect of tBuOOH on lifespan varies between replicates. Several factors may contribute to this, including a variation in temperature of the molten agar and the time spent pouring it, and a variation in duration time required to dry plates. The relative effect of different tBuOOH concentrations on lifespan within each replicate, however, appeared more robust. So, we estimated the relative effects of tBuOOH on lifespan using the multiple regression model

$$\log(y_i) = \beta_c X_i + \beta_r Y_i + \epsilon_j \quad (7)$$

with  $X_i$  as a categorical covariate coding for tBuOOH concentration, and  $Y_i$  as a categorical covariate coding for the biological replicate. The tBuOOH covariate

was coded such that the parameters  $\beta_c$  were relative to the 0 mM category. The replicate name covariate was coded such that the parameters  $\beta_r$  were relative to a single reference replicate. In this way, the parameters  $\beta_c$  represent the best estimate of the relative effect of tBuOOH across replicates.

To determine the quantitative dependence of the scale factor on tBuOOH concentration, the values of  $\beta_c$  and the tBuOOH concentrations corresponding to each  $X_i$  were fitted by the polynomial model  $\lambda = b[\text{tBuOOH}]^a$  using the R<sup>34</sup> nonlinear regression package nls2.

A distinct statistical approach was attempted to confirm the above analysis in a different way. All death times were fitted with the single model

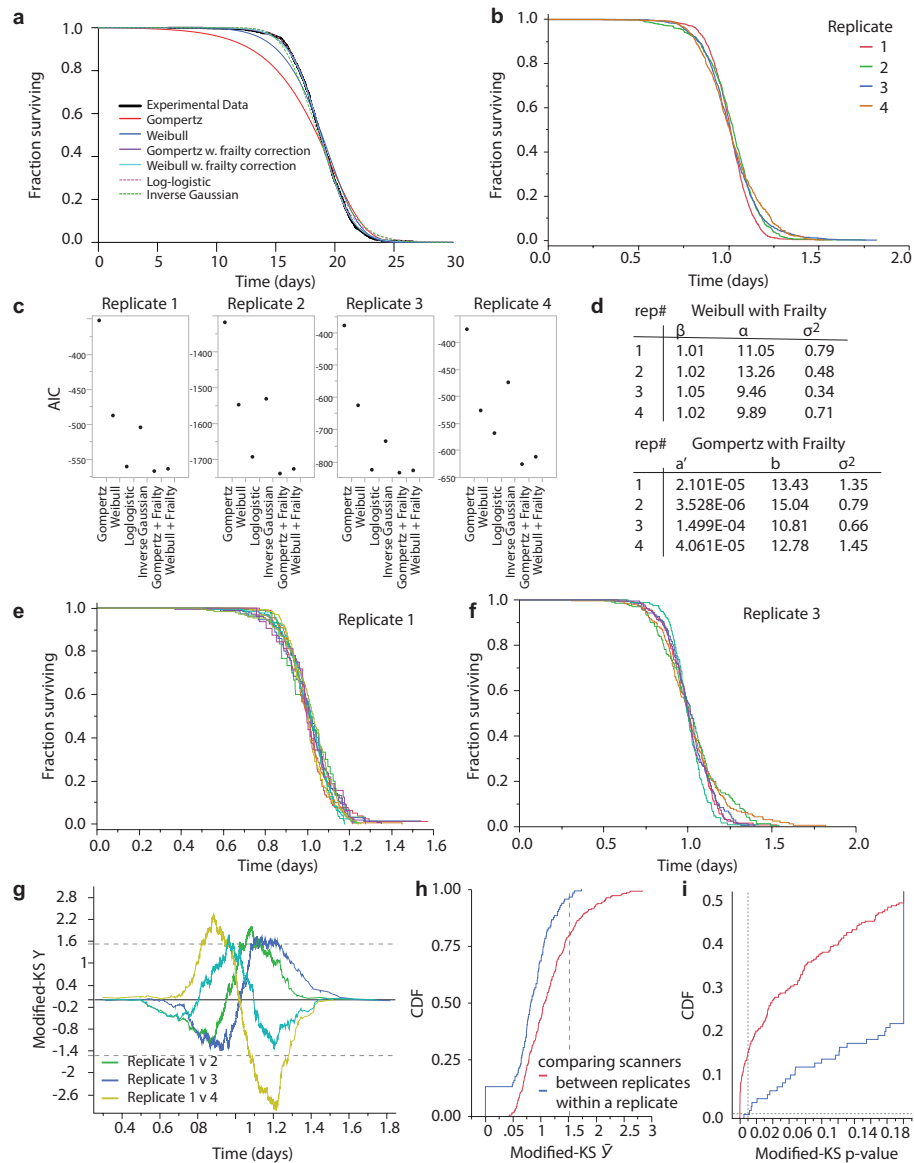
$$\log(y_i) = \beta_c B_i + \beta_r Y_i + \beta_{XB} Y_i B_i + \epsilon_i \quad (8)$$

where  $B_i$  is the logarithm of the tBuOOH concentration to which each individual  $i$  was exposed, represented as a continuous covariate. This makes  $\beta_c$  the best estimate of the exponent of the power-law relationship between lifespan and tBuOOH concentration.  $Y_i$  is the biological replicate in which the individual was observed (animals originating from the same hypochlorite treatment in the same week have the same  $Y_i$ ), and  $Y_i B_i$  is a cross term to identify any systematic differences between the effect of tBuOOH between replicates. The estimated model parameters are provided in Supplementary Table 3.

*Single AFT regression on each tBuOOH replicate (Fig. 4e).* To validate the multiple regression model described previously, the AFT model described in equation (1) was fitted separately on the data collected in each replicate, with  $X_i$  coded to correspond to the tBuOOH concentration. The dummy variables for the tBuOOH covariate were set up such that scale factors were relative to the 0 mM control group. Using these cofactors, the R<sup>34</sup> nonlinear regression package nls2 was used to fit the polynomial relationship  $\lambda = b[\text{tBuOOH}]^a$  for each replicate.

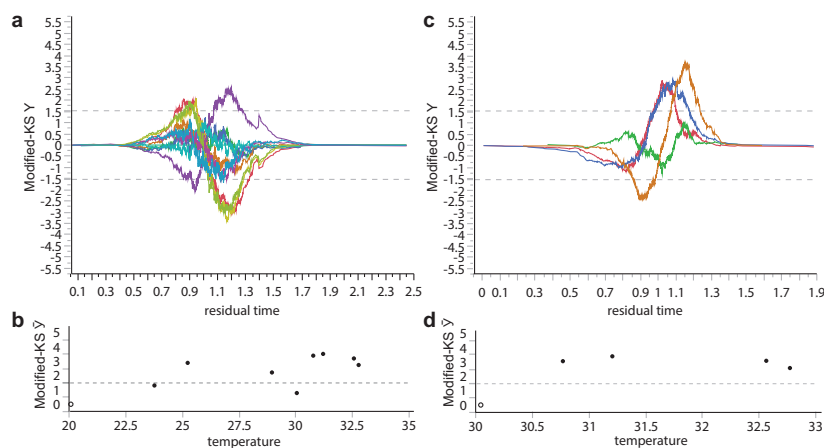
*Estimating the effect of genotype relative to wild type, across tBuOOH concentrations (Fig. 4f).* To measure the effect of mutant alleles on lifespan across t-BuOOH concentrations, a separate AFT model as specified in equation (1) was run on each data set collected at each concentration. In each regression,  $X_i$  was specified as the genotype of individuals, with the categorical variable coded to use wild-type populations as the model intercept.

31. Stiernagle, T. in *WormBook* (ed. The *C. elegans* Research Community) <http://dx.doi.org/10.1895/wormbook.1.101.1> (February 11, 2006).
32. Baeriswyl, S. et al. Modulation of aging profiles in isogenic populations of *Caenorhabditis elegans* by bacteria causing different extrinsic mortality rates. *Biogerontology* **11**, 53–65 (2010).
33. Wilkinson, D. S., Taylor, R. C. & Dillin, A. Analysis of aging in *Caenorhabditis elegans*. *Methods Cell Biol.* **107**, 353–381 (2012).
34. R Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2013).
35. Rebora, P., Salim, A. & Reilly, M. bshazard: a flexible tool for nonparametric smoothing of the hazard function. *R J.* **6**, 114–122 (2014).



**Extended Data Figure 1 | Characterizing the shape of wild-type lifespan distributions at 20°C.** **a**, The AFT residuals corresponding to the 20°C wild-type population presented in Fig. 1 were fitted with a variety of parametric distributions (Supplementary Note 1.4). Fits made to AFT residuals, as opposed to absolute death times, are much less sensitive to any environmental heterogeneity existing between plates and scanners (statistical methods). **b**, The AFT residuals of four additional replicates at 20°C to assess deviations from temporal scaling between replicates. **c**, The Akaike information criterion (AIC) was calculated for each parametric fit of each replicate's AFT residuals shown in **b**. Lower AIC values suggest preferred models. **d**, The parameters of Gompertz and Weibull distributions with frailty corrections are listed; both distributions were good fits across all replicates. **e**, **f**, The survival curves of populations collected in two biological replicates are shown, with one curve for

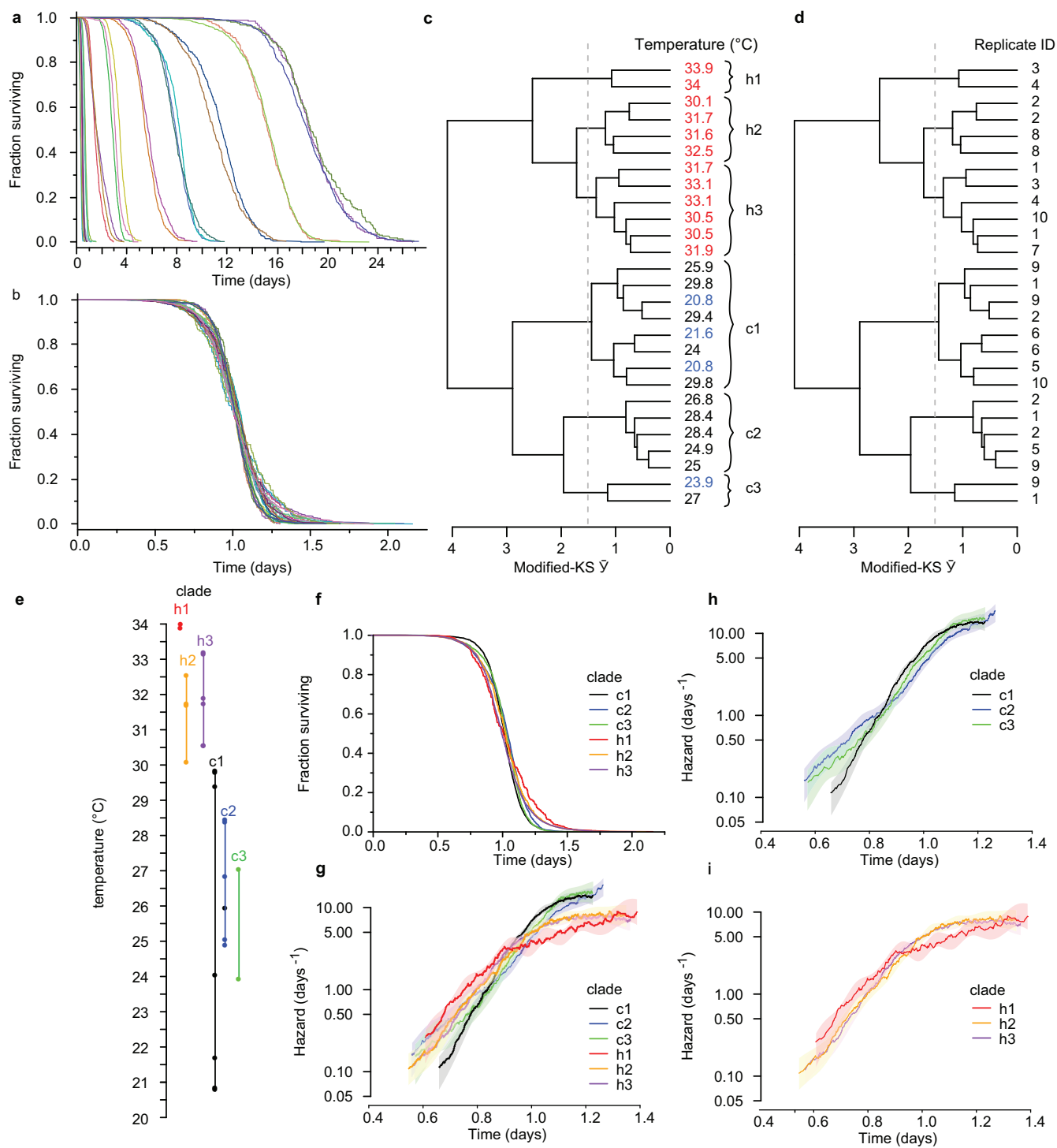
individuals observed on each of 10 and 6 scanners, respectively. **g**, The modified Kolmogorov-Smirnov  $Y(t)$  (Supplementary Note 2) is plotted for comparisons between replicate 1 and all others. Pairs for which  $Y(t) > 1.51$  for some  $t$  exhibit statistically significant deviations from perfect scaling. In this case every replicate differed significantly from the first replicate. **h**, The distribution of modified Kolmogorov-Smirnov test scores,  $\bar{Y}$ , is plotted for comparisons between scanners within a replicate (blue) and between scanners in different replicates (red). Differences between replicates were larger than differences with replicates, suggesting that distance between survival curves observed between scanners cannot alone explain the distance between survival curves observed between replicates. **i**, The  $P$  values corresponding to each  $\bar{Y}$  are shown, with values  $P > 0.01$  considered statistically significant (grey line).



**Extended Data Figure 2 | Apparent deviations from temporal scaling are observed when single replicates are performed at each temperature.**

**a**, For the data shown in Fig. 1a, the modified Kolmogorov–Smirnov score  $Y(t)$  was calculated for AFT residuals (statistical methods and Supplementary Note 2) to compare the reference population at 20 °C with populations held at each of the other temperatures. Pairs for which  $Y(t) > 1.51$  for some  $t$  (grey dashed line) exhibit significant deviations

from perfect scaling. **b**, The modified Kolmogorov–Smirnov test scores  $\bar{Y}$ , corresponding to the maximum absolute value of  $Y(t)$  observed at any time  $t$ , are shown for the comparisons in **a**, highlighting the statistical deviations observed between independent replicates performed at different temperatures. **c**, **d**, The same statistics were calculated when comparing all populations above 30 °C with the population at 30 °C.

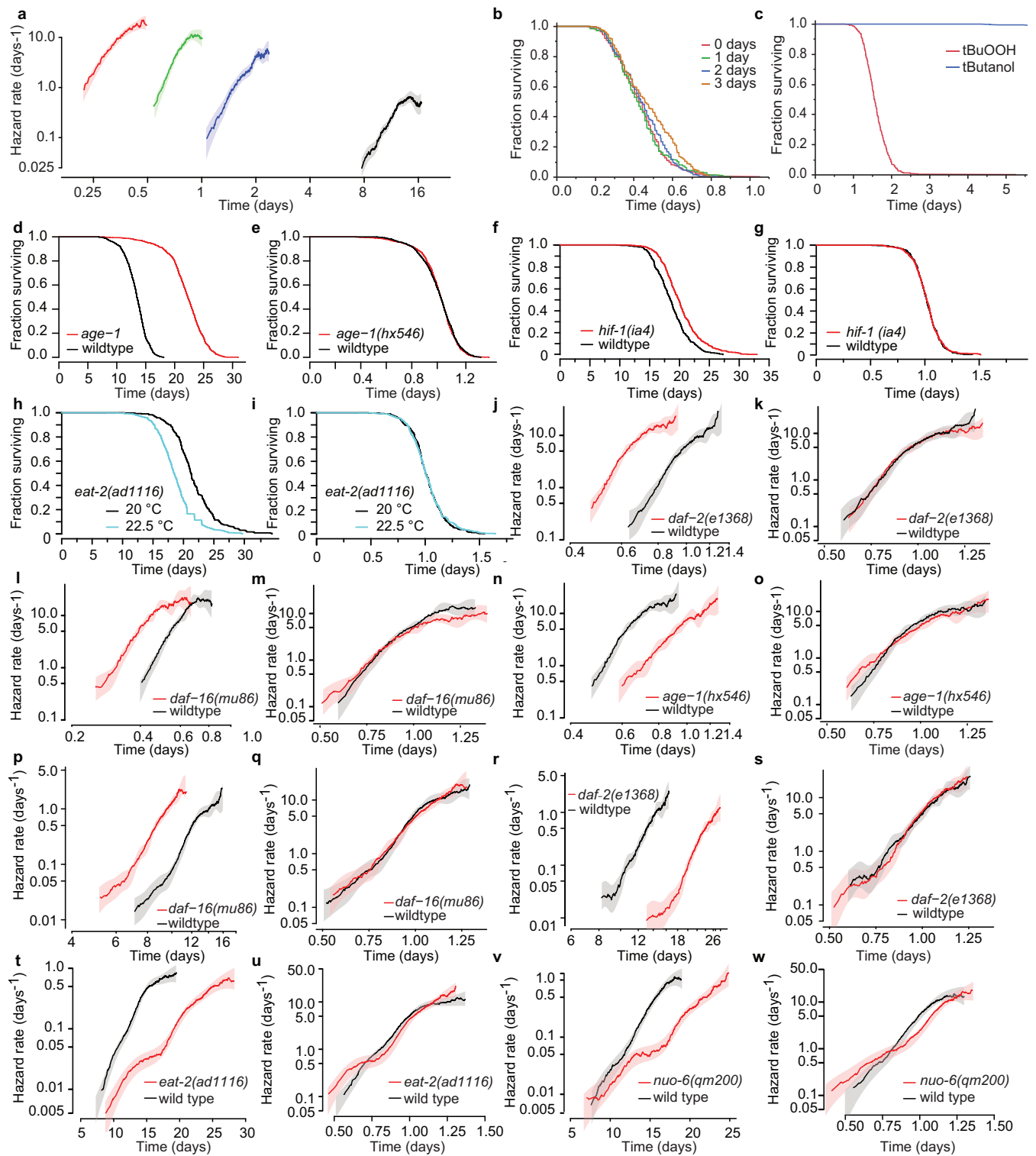


Extended Data Figure 3 | See next page for figure caption.



**Extended Data Figure 3 | Independent replicates demonstrate that apparent deviations from temporal scaling within low- and high-temperature regimes arise from uncontrolled environmental variation.** The lifespan of individuals from populations housed between 20 °C and 34 °C were collected using the lifespan machine (also shown in Fig. 4a, b). To characterize the effects of any uncontrolled experimental conditions specific to individual replicates, and identify any effects of temperature consistent across replicates, we divided the full temperature range into 2 °C intervals. Each 2 °C interval contained lifespan data collected in either two or three independent replicate experiments performed in separate weeks. **a**, Within each 2 °C range, all death times were fitted by an AFT regression model using plate name as a categorical covariate (Statistical methods). The device-corrected death times (the residual time plus model intercept<sup>8</sup>) were plotted, highlighting the changes in survival curve shape between replicates within each 2 °C range. **b**, All deaths across all temperatures were then fitted by a single AFT regression model with plate name as the categorical covariate. AFT residuals were grouped according to their replicate name and temperature range, and plotted to highlight the deviations from temporal scaling across all replicates at all temperatures. **c**, The modified Kolmogorov–Smirnov test (Supplementary Note 2) was applied on each pair of curves shown in **b**. The resulting Kolmogorov–Smirnov  $\bar{Y}$  was used as a distance metric with which to perform a hierarchical clustering, shown as a dendrogram

with each replicate population labelled by the temperature at which it lived. In this dendrogram, populations exhibiting smaller deviations from temporal scaling will have fewer branches between them. Clades that contain statistically significant deviations from temporal scaling have branches extending beyond the dashed grey line, indicating that  $\bar{Y} > 1.51$  between branches. Six statistically distinct groups were identified, three above 30 °C and three below. **d**, The same dendrogram is shown with populations labelled according to the name of the replicate in which they were collected. Populations collected in single replicates did not fall into single clades. This suggests that some environmental factor variable within replicates, distinct from the particular temperature at which populations were placed, produced the observed deviations from temporal scaling. **e**, The statistically distinct clades identified by hierarchical clustering (**c**) are plotted on a temperature scale. Clades overlap at all temperatures except the 30 °C boundary, suggesting that only the 30 °C transition represents a true temperature-dependent deviation from temporal scaling. **f**, The aggregate survival curves containing the AFT residuals of all individuals in each statistically distinct clade are compared, to highlight the differences in shape between clades. **g**, The hazard rate plot of the AFT residuals of all individuals in each statistically distinct clade. **h**, Same as **g**, but showing only the hazard rate plots of populations kept at low temperature. **i**, Same as **g**, but showing only the hazard rate plots of populations kept at high temperature.

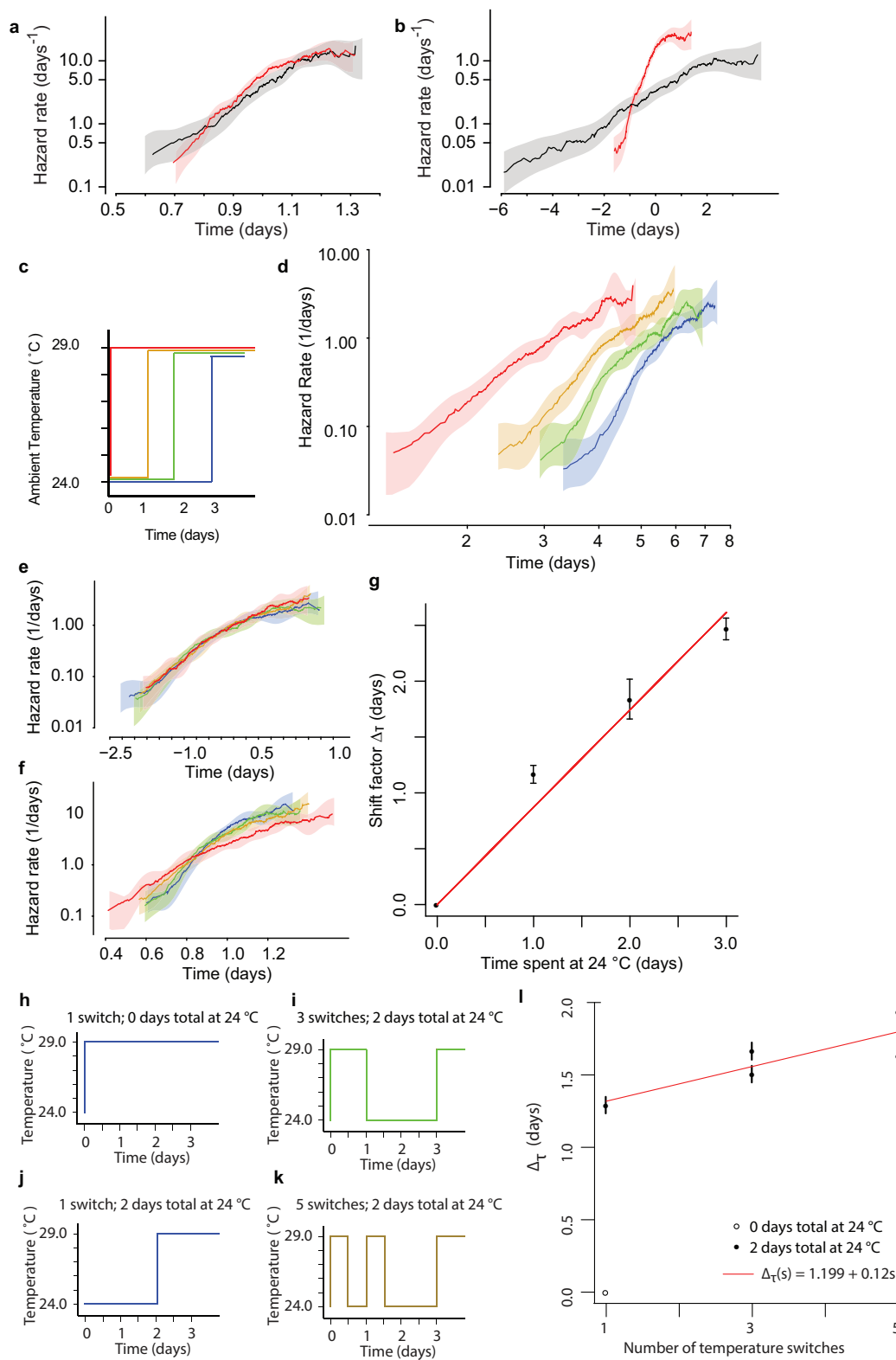


Extended Data Figure 4 | See next page for figure caption.

**Extended Data Figure 4 | Additional survival curves and hazard plots.**

**a**, The hazard rates corresponding to the tBuOOH survival data presented in Fig. 1: 0 mM (black), 1.5 mM (blue), 3 mM (green), and 6 mM (red). **b**, To test for any effects of tBuOOH degradation and evaporation on lifespan, 9 mM tBuOOH plates were prepared and placed at 4 °C. On 4 consecutive days, a subset of plates were seeded with ultraviolet-inactivated bacteria and placed without *C. elegans* on scanners operating at 20 °C. In this way, four groups of plates were created, corresponding to 0, 1, 2, and 3 days of cumulative exposure to standard scanner conditions during which tBuOOH degradation and evaporation could potentially occur. A single age-synchronous population of 2-day-old adult *C. elegans* was then simultaneously distributed across all plates. The remaining lifespan of all worms at 20 °C was recorded using the lifespan machine. **c**, In a separate experiment, plates were prepared containing either 3 mM tBuOOH or 3 mM t-butanol, a degradation product of tBuOOH. On t-butanol, only a trivial fraction of individuals had died by the fifth day, so the experiment was terminated. **d**, **e**, The survival of wild-type and

*age-1(hx546)* populations at 25 °C. **f**, **g**, The lifespan and AFT residuals for *hlf-1(ia4)* and wild-type populations, calculated as in Fig. 2. **h**, **i**, The lifespan and AFT residuals for *eat-2(ad1116)* at 20 °C and 22.5 °C. **j–o**, Age-synchronous mutant (red) and wild-type (black) populations were raised at 25 °C and then transferred to 33 °C on their second day of adulthood, where they remained until death. For each population at 33 °C, the hazard rate was estimated from the death times (**j**, **l**, **m**). The hazard rate was also estimated from the residuals of the AFT regression model  $\log(y_i) = \beta x_i + \epsilon_i$  with genotype as a single categorical covariate (**k**, **m**, **o**). **p**, **q**, The hazard functions of death times and AFT residuals corresponding to the *daf-16(mu86)* data presented in Fig. 1. **r**, **s**, The hazard functions of death times and AFT residuals corresponding to the *daf-2(1368)* data presented in Fig. 1. **t**, **u**, The hazard functions of death times and AFT residuals corresponding to the *eat-2(ad1116)* data presented in Fig. 1. **v**, **w**, The hazard functions of death times and AFT residuals corresponding to the *nuo-6(qm200)* data presented in Fig. 1.

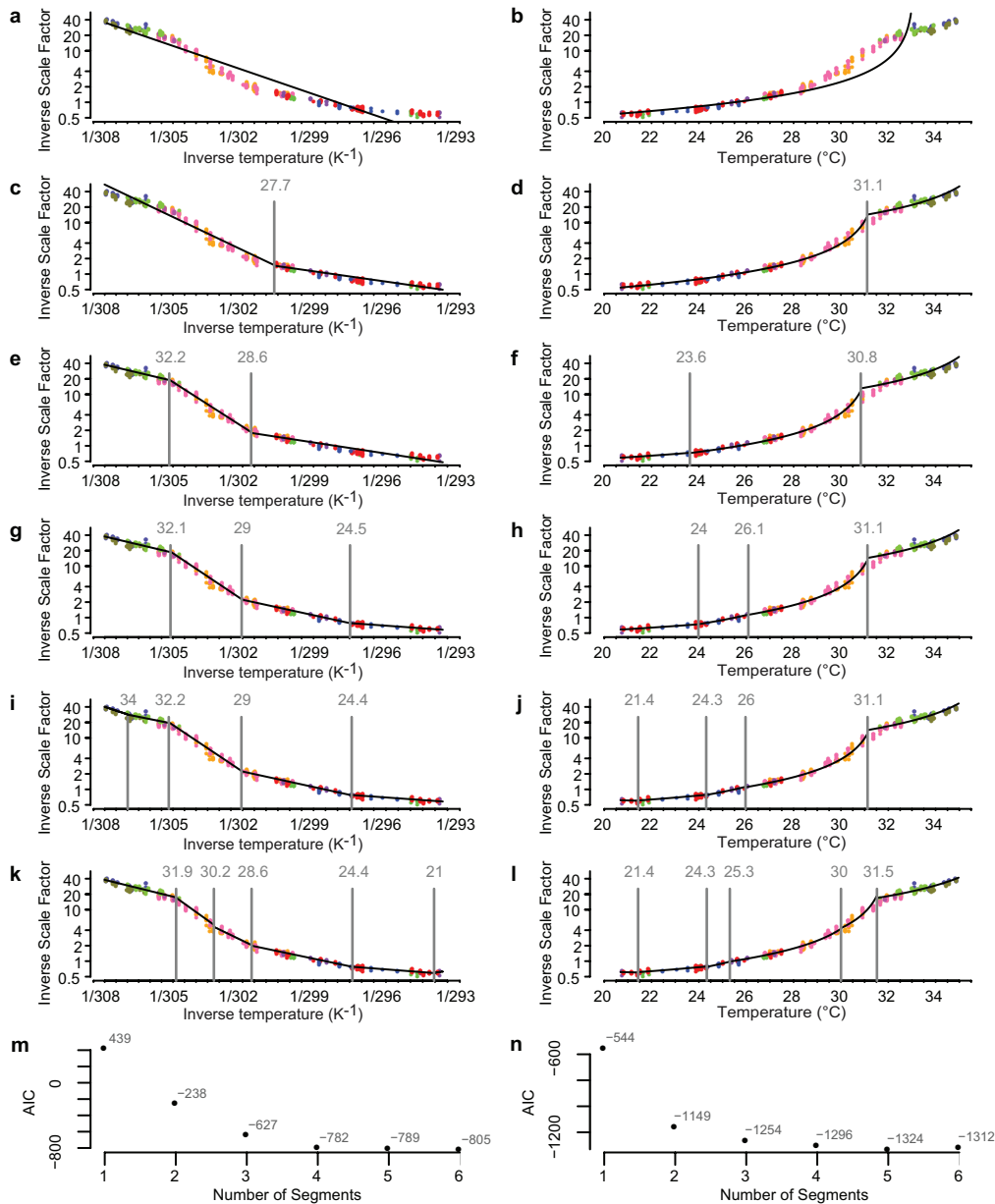


Extended Data Figure 5 | See next page for figure caption.



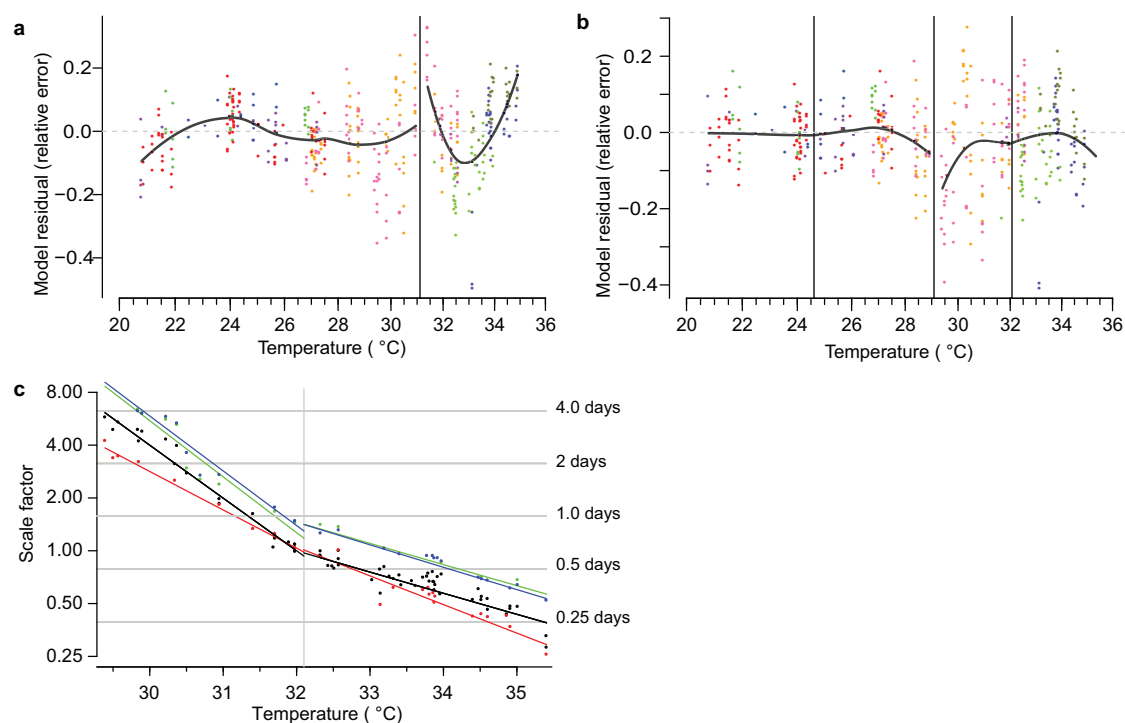
**Extended Data Figure 5 | Additional temperature shift data.** **a**, As a control for the temperature shift experiment shown in Fig. 3, the same regression as in Fig. 3c was run to test for temporal scaling between populations always held at 24 °C (black) and those always held at 29 °C (red). The residuals  $\epsilon_i$  are plotted as hazard functions. **b**, The same regression as in Fig. 3 was run for the same populations as **b** here, to test for temporal shifts. **c**, To test for the effects of different durations spent at 24 °C before transfer to 29 °C. Age-synchronous, wild-type animals were grown at 20 °C and then transferred on their second day of adulthood to 24 °C. Subsets of these animals were then transferred to 29 °C on each of 3 consecutive days. **d**, For each population, the remaining lifespan was observed and the hazard functions estimated. All death times represent the number of days after the second day of adulthood. **e**, The residuals from a regression model with the duration at 24 °C as an additive categorical covariate  $y_i = \beta x_i + \epsilon_i$ . **f**, The residuals from a regression model with the duration at 24 °C as a proportional covariate

$\log(y_i) = \beta x_i + \epsilon_i$ . **g**, The shift values  $\Delta_\tau$  of the additive model are plotted along with a linear fit. **h**, To test for the effect of rapid temperature changes on lifespan, age-synchronous individuals were raised at 20 °C. On their second day of adulthood, a subset was transferred to 29 °C. **i**, Another subset of individuals remained at 24 °C for 2 days, after which they were transferred to 29 °C. **j**, Another subset was transferred to 29 °C, but switched down from 29 °C to 24 °C and then back again to 29 °C (filled circle; three shifts). **k**, A final subset was switched down and back twice (filled circle; five shifts). Note that all populations spent the same total duration at 29 °C, except for the aforementioned control population that was never switched. **l**, The data were fitted with an additive regression model  $y_i = \beta x_i + \epsilon_i$  with the number of switchings as a categorical covariate. The encoding of this covariate was set so that all  $\beta = \Delta_\tau$  represent each subpopulation's change in lifespan relative to the control population that was never switched.



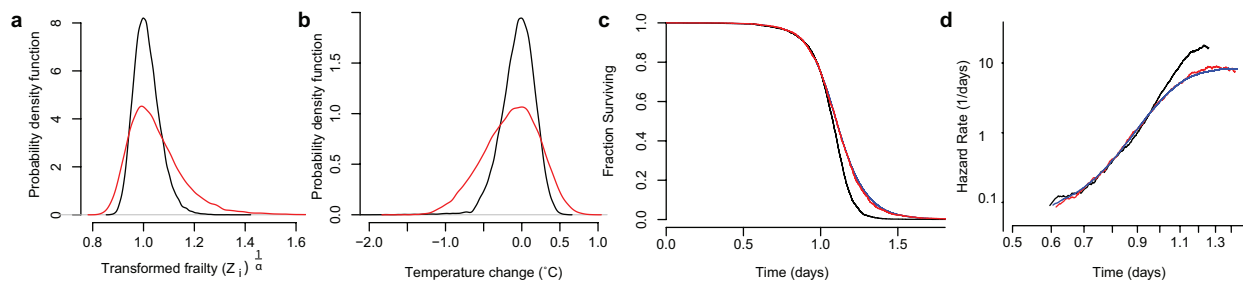
**Extended Data Figure 6 | Identifying the number of thermal regimes, and the boundaries between them.** The thermal scaling data presented in Fig. 4a were fitted with a segmented regression model (statistical methods) assuming that  $\lambda^{-1}$  relates to temperature either following an Arrhenius relationship,  $\lambda(T)^{-1} = p_0 \exp(-p_1/RT)$  (a, c, e, g, i, k) or a linear relationship,  $\lambda(T)^{-1} = p_2 T + p_3$  (b, d, f, h, j, l; statistical methods). As usual we plot the Arrhenius relation on a log-log scale. To emphasize detail, the ordinate of the linear models is also plotted on a logarithmic scale, but the abscissa is kept linear. The model fits are plotted in black, with segment breakpoints shown as vertical lines. Colours correspond to independent biological replicates. Each model was fitted assuming a single segment (a, b), two segments (c, d), three segments (e, f), four segments (g, h), five segments (i, j), or six segments (k, l). m, The AIC corresponding

to each number of segments for the Arrhenius is plotted. n, The AIC for the linear model is plotted. Because the Arrhenius and linear models are fit to distinct data sets (log-transformed and untransformed scale factors respectively, and inverse temperature and untransformed temperatures, respectively), AIC values cannot be compared between Arrhenius and inverse time models. Regime I can be adequately described either by one linear regime or two piecewise Arrhenius regimes, Ia and Ib. Across multiple replicates, the linear model consistently underestimated *C. elegans* lifespan around 25°C, leading us to favour the piecewise Arrhenius model. Regimes II and III involve temperature ranges that are too narrow to distinguish between Arrhenius and linear models. Above 35.5°C, lifespan is too short to be accurately measured with our time-lapse technique.



**Extended Data Figure 7 | Additional data for the regression models in Fig. 4b, d.** The residuals of the linear model (a) shown in Extended Data Fig. 6f. The residuals of the Arrhenius model (b) shown in Fig. 4b and Extended Data Fig. 6g are depicted, showing the deviation of the predicted value from the empirical data across each regime. Residuals are presented in

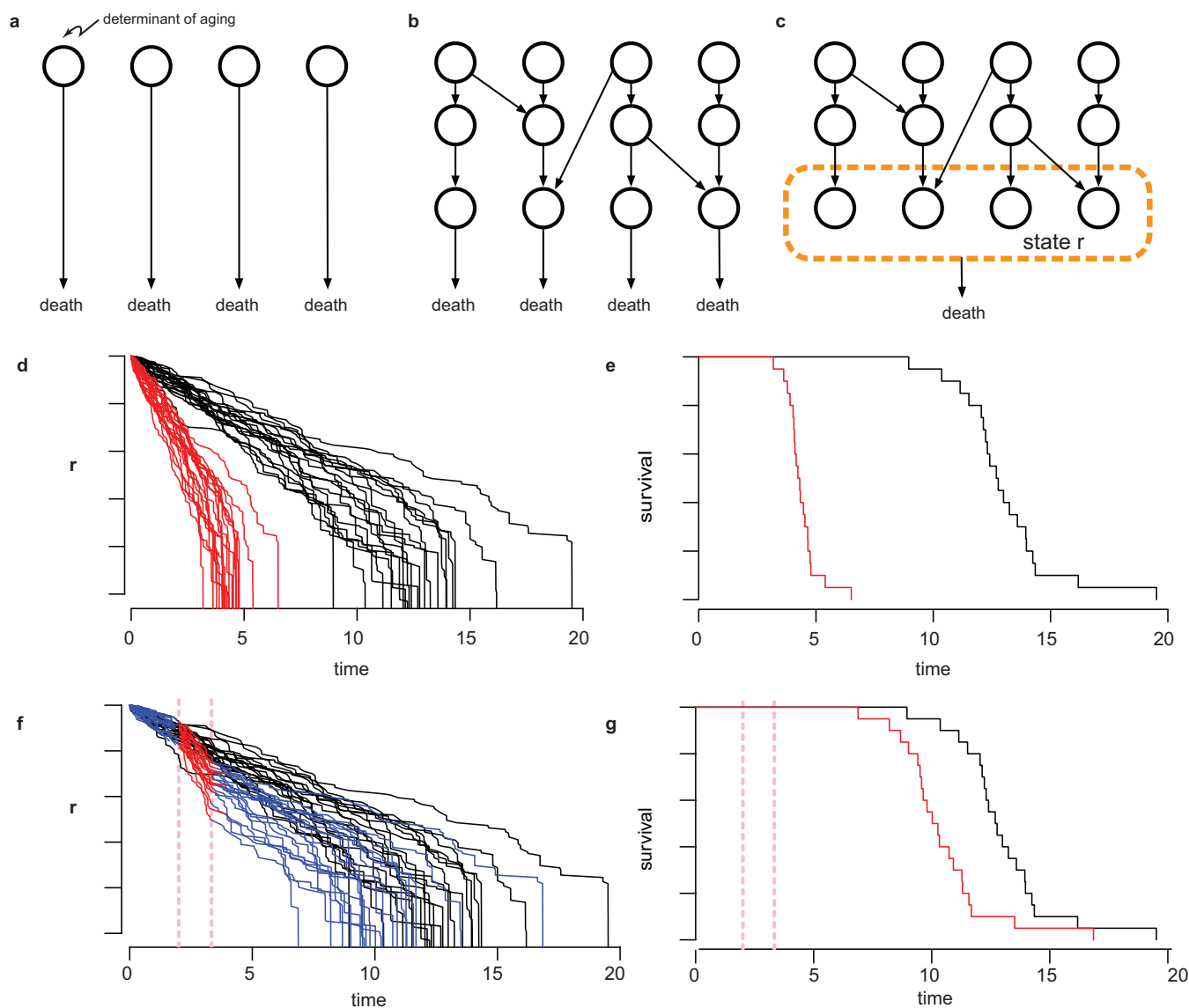
the form of relative error, the ratio between the model's prediction and the empirical measurement. c, As in Fig. 4d, the response of each genotype, wild type (black), *daf-16(mu86)* (red), *daf-2(e1368)* (green), and *age-1(hx546)* (blue) to changes of temperature was estimated (statistical methods) in regime II and in regime III.



**Extended Data Figure 8 | The potential effects of heterogeneity at 33 °C.** **a**, Fifty-six thousand samples were drawn from the distribution of frailty effects  $Z^{-1/\alpha}$  as described in Supplementary Note 3.1, where  $Z$  is a random variable, sampled from an inverse-gamma distribution with mean of 1 and a standard deviation corresponding to the value estimated from experimental data. Samples were drawn using the  $\sigma^2$  estimated for populations at 25 °C (black) and at 33 °C (red), corresponding to the data presented in Fig. 1d. The probability density function of each population is shown, which can be interpreted as the variable effect of unknown factors on lifespan across individuals at each temperature. **b**, At each temperature, 25 °C (black) and at 33 °C (red), we estimated the

distribution of frailties shown in **a**. This was accomplished using the temperature scaling function shown in Fig. 4b. **c**, 56,000 random samples were drawn from the transformed inverse gamma distribution of  $Z^{-1/\alpha}$  with  $\sigma^2$  set to the estimate of  $\Delta\sigma^2$  in equation (15) of Supplementary Note 3.2. Each sample was multiplied by a death time drawn (with replacement) from the set of 25 °C residual times of Fig. 1d, shown here in black. These products constitute a 'transformed' set of death times, corresponding to the 25 °C residuals with additional frailty synthetically introduced. The residual death times of animals placed at 33 °C are shown for comparison (red).





**Extended Data Figure 9 | The organization of lifespan determinants (schematic).** **a**, A set of molecular determinants of risk of death (open circles) do not interact, as is assumed in a competing risks and weakest link models. **b**, Risk determinants might interact (arrows) in complex ways to determine lifespan. In this schematic, each risk of death is still determined by separate factors. **c**, Our data on temporal scaling suggest that the set of molecular determinants that determine risk of death (within the dotted circle) must change in concert when exposed to interventions in ageing. This set is therefore well described by a single state variable  $r$ .

**d**, A cartoon of the stochastic decline of such a state variable (generated from the dependency model discussed in Supplementary Note 5.2). Each trajectory represents the values of  $r$  over time for each individual in a population. Interventions affect the dynamics of the state decline by rescaling the average dynamics of exposed individuals (red lines), which produces (e) a rescaling of the resultant survival curve. **f**, Transient interventions in young adults (applied within the red dotted vertical lines) transiently rescale the average dynamics, leading to (g) a shift in the lifespan distribution.

# An essential receptor for adeno-associated virus infection

S. Pillay<sup>1\*</sup>, N. L. Meyer<sup>2\*</sup>, A. S. Puschnik<sup>1</sup>, O. Davulcu<sup>2</sup>, J. Diep<sup>1</sup>, Y. Ishikawa<sup>2,3</sup>, L. T. Jae<sup>4</sup>, J. E. Wosen<sup>1</sup>, C. M. Nagamine<sup>5</sup>, M. S. Chapman<sup>2</sup> & J. E. Carette<sup>1</sup>

**Adeno-associated virus (AAV) vectors are currently the leading candidates for virus-based gene therapies because of their broad tissue tropism, non-pathogenic nature and low immunogenicity<sup>1</sup>. They have been successfully used in clinical trials to treat hereditary diseases such as haemophilia B (ref. 2), and have been approved for treatment of lipoprotein lipase deficiency in Europe<sup>3</sup>. Considerable efforts have been made to engineer AAV variants with novel and biomedically valuable cell tropisms to allow efficacious systemic administration<sup>1,4</sup>, yet basic aspects of AAV cellular entry are still poorly understood. In particular, the protein receptor(s) required for AAV entry after cell attachment remains unknown. Here we use an unbiased genetic screen to identify proteins essential for AAV serotype 2 (AAV2) infection in a haploid human cell line. The most significantly enriched gene of the screen encodes a previously uncharacterized type I transmembrane protein, KIAA0319L (denoted hereafter as AAV receptor (AAVR)). We characterize AAVR as a protein capable of rapid endocytosis from the plasma membrane and trafficking to the *trans*-Golgi network. We show that AAVR directly binds to AAV2 particles, and that anti-AAVR antibodies efficiently block AAV2 infection. Moreover, genetic ablation of AAVR renders a wide range of mammalian cell types highly resistant to AAV2 infection. Notably, AAVR serves as a critical host factor for all tested AAV serotypes. The importance of AAVR for *in vivo* gene delivery is further highlighted by the robust resistance of *Aavr*<sup>-/-</sup> (also known as *Au040320*<sup>-/-</sup> and *Kiaa0319l*<sup>-/-</sup>) mice to AAV infection. Collectively, our data indicate that AAVR is a universal receptor involved in AAV infection.**

AAV2, the most commonly studied AAV serotype, attaches to cells using heparan sulfate proteoglycan<sup>5</sup>. For several other non-enveloped viruses, initial attachment is followed by engagement of a protein receptor, which dictates entry into the cytoplasm. Whether AAV also requires such a protein receptor is unclear. Surface proteins including human fibroblast growth factor receptor-1 (FGFR1) and hepatocyte growth factor receptor (MET) have been reported as putative AAV2 co-receptors<sup>6,7</sup>. However, using isogenic knockout cell lines (Extended Data Fig. 1a, b), we observed no significant effect on AAV2 infection in cells lacking FGFR1, and only minimal effect as a result of MET loss (Extended Data Fig. 1c), suggesting a modest role in AAV2 infection for these proteins. To identify host factors critical for AAV2 infection, we used an unbiased genome-wide screening approach based on insertional mutagenesis in haploid human cells (HAP1)<sup>8</sup>. We infected a library of mutagenized cells, carrying knockouts in virtually all non-essential genes, with an AAV2 vector that expresses red fluorescent protein (RFP) (Extended Data Fig. 2a). Mutant cells refractory to AAV2 infection were isolated through iterative cycles of fluorescence-activated cell sorting (FACS) (Extended Data Fig. 2b). The screen yielded 46 significant hits (Fig. 1a, Supplementary Table 1), many of

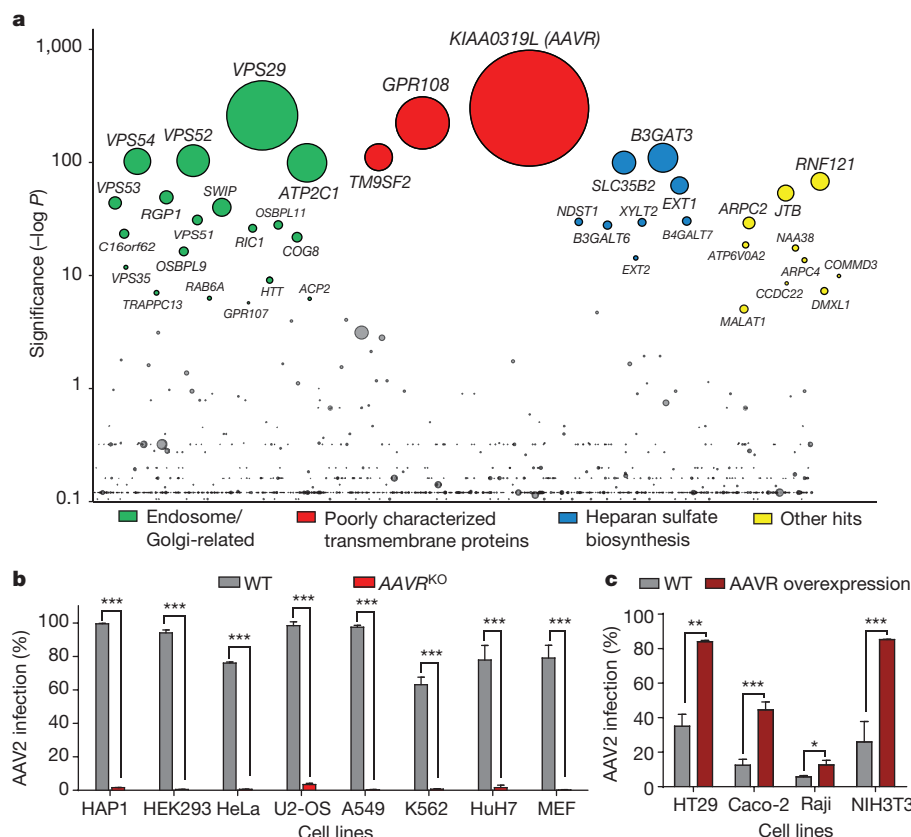
which were implicated in heparan sulfate proteoglycan biosynthesis (depicted in blue). AAV2 hijacks endosomal pathways to travel from the cell surface to the nucleus, and several endosomal trafficking genes (depicted in green) were prominently identified in the screen, specifically those encoding members of the retromer (VPS29, VPS35) and Golgi-associated retrograde protein complexes (VPS51, VPS52, VPS53, VPS54). These proteins are involved in retrograde transport from the endosomes to the Golgi<sup>9,10</sup>, but have not been specifically associated with AAV2 infection before now. The most significantly enriched gene of the screen was *KIAA0319L* (denoted hereafter as AAVR), with 570 independent mutations identified. This gene encodes a poorly characterized transmembrane protein. Little is known about the cellular function of AAVR, but it has been linked to dyslexia, with a potential role in neuronal migration<sup>11</sup>.

To validate the role of AAVR in AAV2 infection, we used CRISPR-Cas9 genome engineering to generate isogenic AAVR-knockout (AAVR<sup>KO</sup>) cell lines in a panel of cell types representing various human and mouse tissues (Extended Data Table 1). In all eight cell types, AAVR knockout rendered cells highly resistant to AAV2 infection (20,000 viral genomes (vg) per cell) (Fig. 1b). At a multiplicity of infection as high as 100,000 vg per cell, AAVR<sup>KO</sup> cells still remained poorly susceptible to infection by an AAV2-luciferase vector (Extended Data Fig. 3a). This also held true for wild-type AAV2 infection, where AAV2 replication was negligible in AAVR<sup>KO</sup> cells (Extended Data Fig. 3b). Notably, *MET* and *FGFR1* knockouts demonstrated no resistance to infection in multiple cell types (Extended Data Fig. 3e). Genetic complementation of AAVR in AAVR<sup>KO</sup> cells (Extended Data Fig. 3c) restored susceptibility to AAV2 in all cell types assessed, confirming that the resistance phenotype observed in AAVR<sup>KO</sup> cells was caused by loss of AAVR expression (Extended Data Fig. 3d). To examine further if AAVR expression can limit AAV2 infection, we overexpressed AAVR in four cell lines previously identified as poorly permissive to AAV2 (refs 12, 13). We observed an increase in susceptibility to AAV2 in all AAVR-overexpressing cell lines compared to wild-type cells, emphasizing the important role of AAVR in AAV2 infection (Fig. 1c).

AAVR is a predicted type I transmembrane protein with five immunoglobulin-like (Ig-like) domains in its ectodomain, referred to as polycystic kidney disease (PKD) domains<sup>14</sup>. Ig-like domains mediate cell-cell adhesion and are present in various well-characterized virus receptors<sup>15</sup>, including those for poliovirus, measles virus and reovirus. On the basis of their similarity to other receptors and the heavy dependence of AAV2 infection on AAVR, we hypothesized that AAVR acts as an AAV2 receptor. We first determined whether AAVR PKD domains are responsible for mediating AAV2 infection by creating a series of AAVR deletion mutants and expressing each in AAVR<sup>KO</sup> cells (Fig. 2a). Simultaneous deletion of AAVR PKD domains 1 and 2, or 2 and 3, abrogated its role in AAV2 infection, whereas deletions

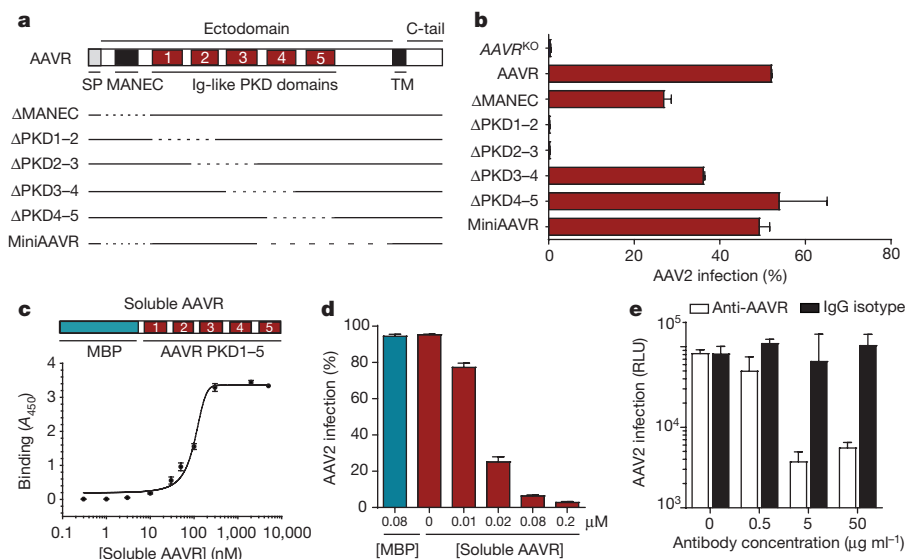
<sup>1</sup>Department of Microbiology and Immunology, Stanford University School of Medicine, 299 Campus Drive, Stanford, California 94305, USA. <sup>2</sup>Department of Biochemistry and Molecular Biology, School of Medicine, Oregon Health & Science University, 3181 Sam Jackson Park Road, Portland, Oregon 97239-3098, USA. <sup>3</sup>Shriners Hospital for Children, 3101 Sam Jackson Park Road, Portland, Oregon 97239, USA. <sup>4</sup>Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, Netherlands. <sup>5</sup>Department of Comparative Medicine, Stanford University School of Medicine, 287 Campus Drive, Stanford, California 94305, USA.

\*These authors contributed equally to this work.



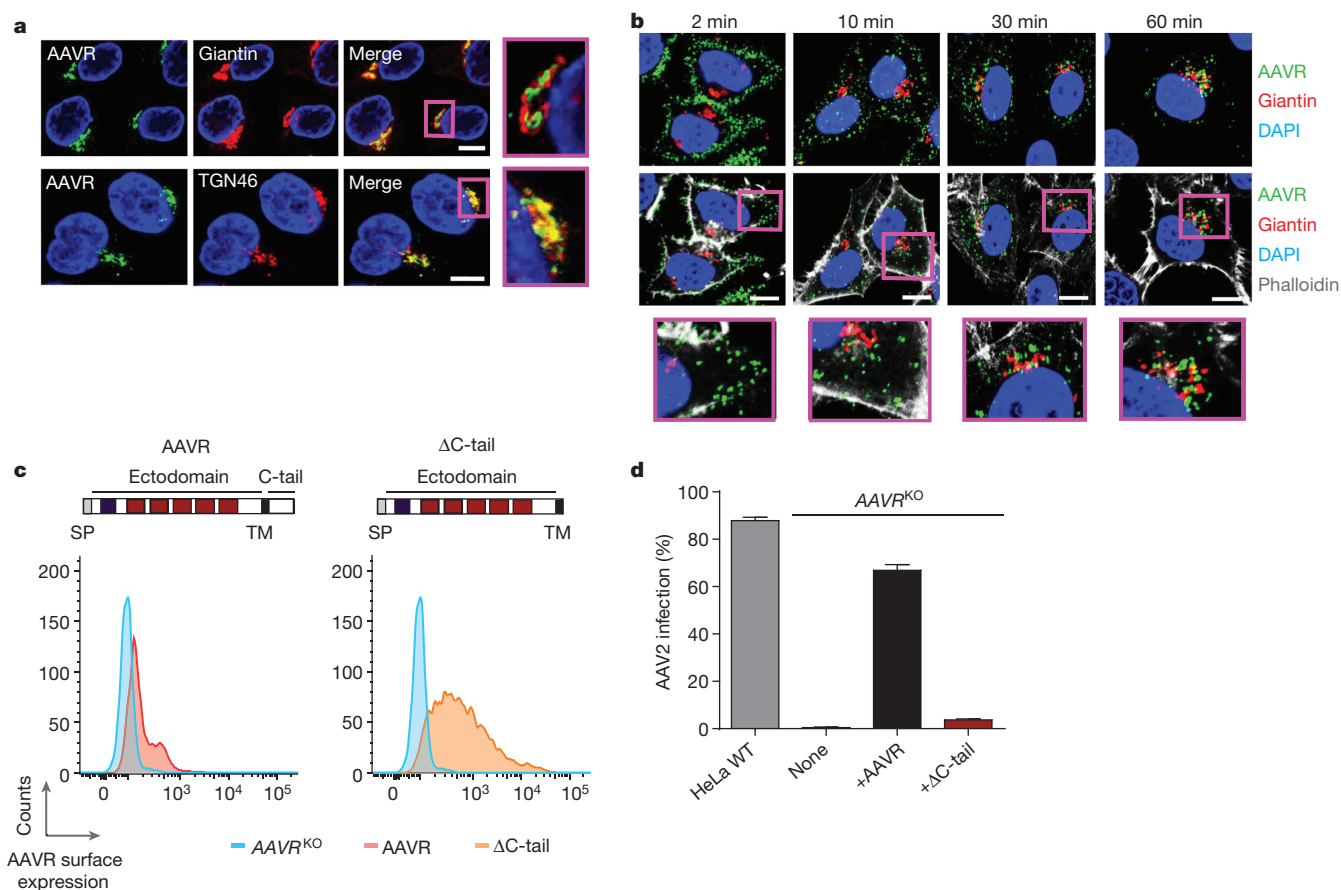
**Figure 1 | An unbiased haploid genetic screen identifies KIAA0319L (AAVR), an essential host factor for AAV2 infection.** **a**, Bubble plot illustrating the significance of enrichment of gene-trap insertions within identified genes (relative to unselected control population). Bubbles represent genes with width proportional to number of independent gene-trap insertions. Top 40 significant genes ( $P \leq 0.001$ ) are coloured and grouped by function. **b**, AAV2-RFP infection in wild-type (WT) cells and

AAVR knockout (AAVR<sup>KO</sup>) cells, evaluated in AAV2-susceptible human and mouse cell lines. **c**, AAV2-RFP infection of poorly permissive human and mouse cell lines with and without AAVR overexpression. Data depict mean with s.d. for triplicate infections. Infections were performed using a multiplicity of infection (MOI) of 20,000 vg per cell for 24 h. Significance was determined using unpaired parametric two-sided Student's *t*-test, with a Welch post-correction; \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .



**Figure 2 | AAVR binds specifically to AAV2 via Ig-like PKD domains.** **a**, Schematic of AAVR domains and deletion mutants; dotted lines represent deletions. **b**, AAV2-RFP infection of HAP1 AAVR<sup>KO</sup> cells expressing AAVR deletion mutants (MOI: 20,000 vg per cell). **c**, ELISA showing binding of soluble AAVR (fusion protein between MBP and AAVR PKD1-5) to AAV2 particles. **d**, AAV2 neutralization assay of cells incubated with soluble AAVR or MBP during AAV2-GFP infection

(MOI: 7,500 vg per cell). **e**, Antibody inhibition assay of wild-type HeLa cells incubated with anti-AAVR or IgG isotype control antibodies (at indicated concentrations) at 4 °C before AAV2-luciferase infection (MOI: 1,000 vg per cell). Data depict mean with s.d. for triplicate infections; transgene expression measured after 24 h. MANEC, motif at amino terminus with eight cysteines; RLU, relative light units; SP, signal peptide; TM, transmembrane.



**Figure 3 | AAVR traffics from the plasma membrane to the *trans*-Golgi network and AAVR endocytosis is necessary for AAV2 infection.**

**a**, Endogenous AAVR localization in wild-type HeLa cells shown with markers for *cis*-medial Golgi (giantin) and *trans*-Golgi network (TGN46). **b**, Tracking AAVR endocytosis using anti-AAVR antibodies. AAVR-complement cells were incubated with anti-AAVR antibodies for 1 h at 4 °C, washed and then incubated at 37 °C. At respective time points,

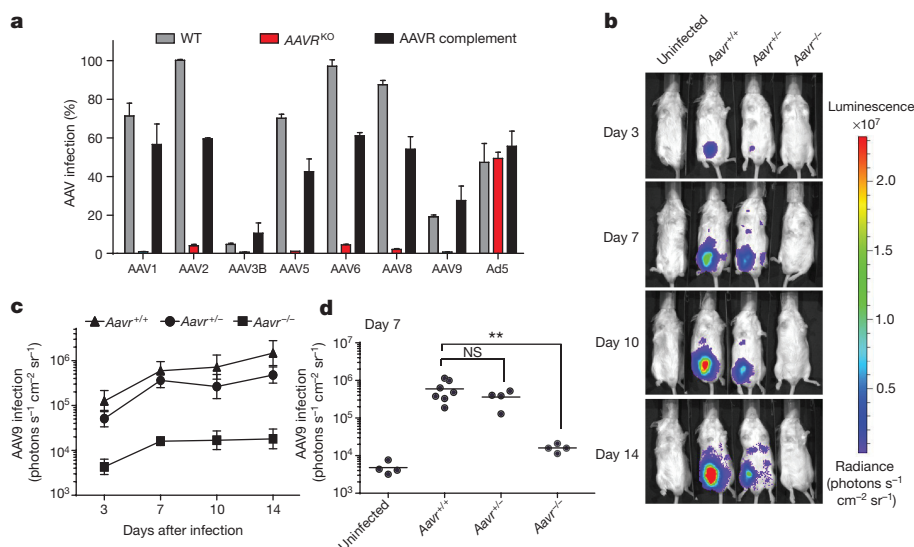
in other regions were tolerated (Fig. 2b). An AAVR minimal mutant (miniAAVR) consisting of PKD domains 1–3 in the ectodomain efficiently rescued AAV2 infection, highlighting the importance of the first three PKD domains for infection. Notably, soluble AAVR (recombinant protein comprising a fusion between maltose-binding protein (MBP) and AAVR PKD domains 1–5 expressed in *Escherichia coli*), but not MBP alone, bound directly to AAV2 particles (Fig. 2c and Extended Data Fig. 4a) with a dissociation constant ( $K_d$ ) of ~150 nM (measured using surface plasmon resonance, see Extended Data Fig. 4b). We next investigated whether AAV2 infection could be neutralized in the presence of soluble AAVR. Indeed, infection efficiency was inhibited in a concentration-dependent manner when soluble AAVR was included during infection (Fig. 2d and Extended Data Fig. 4c). Consistent with this inhibition assay, antibodies directed against AAVR were capable of potently blocking AAV2 infection by more than tenfold when incubated with cells before infection, in contrast to control IgG antibodies (Fig. 2e). This suggests that blocking viral access to AAVR on the cell surface substantially limits infection.

Characterization of the subcellular localization of AAVR revealed distinct perinuclear localization, demonstrating a strong association with the *cis*-medial Golgi marker (giantin), and complete co-localization with the *trans*-Golgi network (TGN) marker (TGN46) (Fig. 3a). Many TGN proteins are dynamically recycled from the plasma membrane through motifs in their carboxy terminal cytoplasmic tail (C-tail) that direct endocytosis and intracellular trafficking (reviewed in refs 16, 17). To determine whether AAVR is such a recycling receptor, we specifically labelled the cell surface pool of AAVR by incubating live

cells were fixed and anti-AAVR antibodies were visualized to track the trafficking of surface AAVR. **c**, AAVR surface expression in AAVR<sup>KO</sup> cells with and without overexpression of full-length AAVR and ΔC-tail (depicted in schematic). **d**, AAV2-RFP infection (MOI: 20,000 vg per cell; measured after 24 h) in AAVR<sup>KO</sup> cells stably expressing constructs depicted in **c**. Data depict the mean with s.d. for triplicate infections. Scale bars, 10 μm.

AAVR-complement cells with anti-AAVR antibodies under cold conditions. Cells were warmed to initiate endocytosis and fixed at defined time points. Labelled AAVR gradually moved from the surface into the cell, and concentrated in a perinuclear location associated with the Golgi marker (Fig. 3b). This rapid endocytosis may explain why we did not observe AAVR at the cell surface in steady state (Fig. 3a). As a control, AAVR<sup>KO</sup> cells were labelled similarly to AAVR-expressing cells, but no AAVR was detected on these cells (Extended Data Fig. 5a). Interestingly, the intracellular trafficking route of AAVR mapped here is remarkably similar to that of AAV particles, trafficking from the plasma membrane to the Golgi<sup>18</sup>. To determine whether AAVR endocytosis contributes to mediation of AAV2 infection, we removed the C-tail, which encodes endocytic motifs of AAVR. C-tail deletion (ΔC-tail) led to increased cell surface expression of AAVR (Fig. 3c and Extended Data Fig. 5b) and prevented endocytosis (Extended Data Fig. 5c). Importantly, the ΔC-tail was incapable of mediating AAV2 infection upon complementation in AAVR<sup>KO</sup> cells (Fig. 3d), suggesting that AAVR endocytosis is required for AAV2 infection. We further investigated whether AAVR requires intracellular trafficking to the TGN to mediate infection by replacing the C-tail of miniAAVR with those of cellular receptors with well-characterized endocytic motifs (Extended Data Fig. 6a). These included the cation-independent mannose 6-phosphate receptor (Ci-MPR), the prototypical receptor that mediates trafficking from the plasma membrane to the TGN<sup>19</sup>. We also included the low density lipoprotein receptor (LDLR)<sup>20</sup> and poliovirus receptor (PVR)<sup>21</sup>, both of which endocytose and traffic between the plasma membrane and endosomes but are not reported to travel to





**Figure 4 | AAVR is a critical host factor for the infection of naturally occurring AAV serotypes and essential for AAV infection *in vivo*.**

**a**, Infection of wild-type HeLa cells, AAVR<sup>KO</sup> cells and AAVR<sup>KO</sup> cells overexpressing AAVR (AAVR complement), using AAV vectors of different serotypes (MOI:  $10^5$  vg per cell; RFP/GFP expression measured at 24 h). **b**, Bioluminescence of AAV9-infected wild-type (Aavr<sup>+/+</sup>), heterozygous (Aavr<sup>+/-</sup>) and Aavr<sup>KO</sup> (Aavr<sup>-/-</sup>) FVB mice over 14 days;

representative images of mice from each group are shown with a radiance range of  $5 \times 10^5$  to  $1 \times 10^7$  photons  $s^{-1} cm^{-2} sr^{-1}$ . **c**, AAV9-luciferase infection for Aavr<sup>+/+</sup>, Aavr<sup>+/-</sup> and Aavr<sup>-/-</sup> groups (measured as average radiance) at the respective days after infection. **d**, AAV9-luciferase infection of mice at day 7. Data depicts the mean (with s.d. in **a** and **c**). Significance was determined using unpaired two-sided Mann–Whitney *t*-test; \*\**P* < 0.01; NS, not significant.

the TGN. Each of the fusion constructs displayed cellular localization patterns comparable to their parent receptors, as PVR-tail and LDLR-tail were detectable on the cell surface, and Ci-MPR-tail co-localized with a TGN marker and displayed a dispersed pattern in the cytoplasm (Extended Data Fig. 6b). Remarkably, all fusion constructs rescued AAV2 infection, albeit to different degrees (Extended Data Fig. 6c). Compared to the restored infection in miniAAVR-expressing cells, the LDLR and PVR fusion constructs yielded twofold and fourfold reduced infectivity, respectively. Conversely, routing AAVR to the TGN using Ci-MPR endocytosis signals resulted in infection rates similar to miniAAVR-expressing cells. Thus, trafficking to the TGN appears to increase AAV infection efficiency, but varying rates of endocytosis between the constructs may also contribute to infectivity differences. Nonetheless, all fusion constructs rescued infectivity to an extent, even those that are not reported to traffic to the TGN, indicating that trafficking to the TGN may not be a strict requirement. Collectively, these data suggest a model where AAVR interacts with AAV at the cell surface and facilitates trafficking to the TGN, but do not exclude the possibility that the interaction is initiated in early/late endosomes and/or the TGN.

To test whether other naturally occurring AAV serotypes are also dependent on AAVR, we infected AAVR<sup>KO</sup> cells with a panel of AAV serotypes including AAV1, 2, 3B, 5, 6, 8 and 9 (expressing green fluorescent protein (GFP) or RFP). We also infected cells with a recombinant adenovirus 5 vector expressing RFP (Ad5-RFP). AAVR<sup>KO</sup> cells displayed a robust resistance to all AAV serotypes (Fig. 4a), irrespective of the different glycan attachment factors used by each serotype. AAV susceptibility was also restored in AAVR-complemented cells, as previously observed with AAV2. Moreover, there was no significant difference in Ad5-RFP infection among the three cell lines tested. The role of AAVR in infection for the tested viruses is therefore specific to AAV and is ubiquitously required for a variety of human- and simian-derived AAV serotypes.

Finally, we tested the contribution of AAVR to *in vivo* gene delivery. We generated Aavr-knockout mice (Aavr<sup>-/-</sup>) using transcription activator-like effector nuclease (TALEN)-mediated gene targeting. Aavr<sup>-/-</sup> mice did not display any apparent developmental or physical phenotype. Wild-type (Aavr<sup>+/+</sup>), heterozygous (Aavr<sup>+/-</sup>) and Aavr<sup>-/-</sup> FVB mice (genotypes depicted in Extended Data Fig. 7a)

were injected intraperitoneally with AAV9-luciferase, chosen because of its high transduction efficiency *in vivo* compared to AAV2 (ref. 22). Bioluminescence (a measure of luciferase expression) was strongest in the lower abdomen of Aavr<sup>+/+</sup> mice, intensifying over 14 days (Fig. 4b, c and Extended Data Fig. 7b). Aavr heterozygosity did not significantly reduce AAV9 infection *in vivo*; however, Aavr<sup>-/-</sup> mice displayed a pronounced reduction in bioluminescence, comparable to background levels obtained in uninfected wild-type mice (Fig. 4d and Extended Data Fig. 7c).

Overall, this study identifies AAVR as a key host receptor for AAV infection *in vitro* and *in vivo*, using an unbiased and comprehensive genetic screening approach. AAV vector usage for gene therapy is rapidly growing, and recent advances in genome editing<sup>23</sup> and vectored immunoprophylaxis<sup>24</sup> are expected to further expand its utility. Exploiting AAVR as a tool to improve AAV-based applications may enhance its efficacy in basic research and clinical settings. An understanding of AAVR tissue distribution will be important in determining the contribution of AAVR to AAV tropism, although additional factors<sup>25</sup> including glycan usage, AAV nuclear import and other genes identified in our screen could also influence tropism. Lastly, AAV vectors are commonly used in experimental mouse models; hence expression of AAVR under specific promoters (for example, for cells in the substantia nigra) in an Aavr<sup>-/-</sup> background may aid in developing better mouse models for human diseases such as those for neurological disorders.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 29 May; accepted 18 November 2015.**

**Published online 27 January 2016.**

- Kotterman, M. A. & Schaffer, D. V. Engineering adeno-associated viruses for clinical gene therapy. *Nature Rev. Genet.* **15**, 445–451 (2014).
- Nathwani, A. C. *et al.* Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N. Engl. J. Med.* **365**, 2357–2365 (2011).
- Gaudet, D., Methot, J. & Kastelein, J. Gene therapy for lipoprotein lipase deficiency. *Curr. Opin. Lipidol.* **23**, 310–320 (2012).
- Lisowski, L. *et al.* Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. *Nature* **506**, 382–386 (2014).
- Summerford, C. & Samulski, R. J. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J. Virol.* **72**, 1438–1445 (1998).

6. Kashiwakura, Y. *et al.* Hepatocyte growth factor receptor is a coreceptor for adeno-associated virus type 2 infection. *J. Virol.* **79**, 609–614 (2005).
7. Qing, K. *et al.* Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. *Nature Med.* **5**, 71–77 (1999).
8. Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477**, 340–343 (2011).
9. Bonifacino, J. S. & Hierro, A. Transport according to GARP: receiving retrograde cargo at the *trans*-Golgi network. *Trends Cell Biol.* **21**, 159–167 (2011).
10. McGough, I. J. & Cullen, P. J. Recent advances in retromer biology. *Traffic* **12**, 963–971 (2011).
11. Poelmans, G., Buitelaar, J. K., Pauls, D. L. & Franke, B. A theoretical molecular network for dyslexia: integrating available genetic findings. *Mol. Psychiatry* **16**, 365–382 (2011).
12. Ellis, B. L. *et al.* A survey of *ex vivo/in vitro* transduction efficiency of mammalian primary cells and cell lines with nine natural adeno-associated virus (AAV1–9) and one engineered adeno-associated virus serotype. *Virol. J.* **10**, 74 (2013).
13. Hansen, J., Qing, K., Kwon, H. J., Mah, C. & Srivastava, A. Impaired intracellular trafficking of adeno-associated virus type 2 vectors limits efficient transduction of murine fibroblasts. *J. Virol.* **74**, 992–996 (2000).
14. Ibraghimov-Beskrovnaya, O. *et al.* Strong homophilic interactions of the Ig-like domains of polycystin-1, the protein product of an autosomal dominant polycystic kidney disease gene, PKD1. *Hum. Mol. Genet.* **9**, 1641–1649 (2000).
15. Bhella, D. The role of cellular adhesion molecules in virus attachment and entry. *Phil. Trans. R. Soc. Lond. B* **370**, 20140035 (2015).
16. Maxfield, F. R. & McGraw, T. E. Endocytic recycling. *Nature Rev. Mol. Cell Biol.* **5**, 121–132 (2004).
17. Kelly, B. T. & Owen, D. J. Endocytic sorting of transmembrane protein cargo. *Curr. Opin. Cell Biol.* **23**, 404–412 (2011).
18. Nonnenmacher, M. & Weber, T. Adeno-associated virus 2 infection requires endocytosis through the CLIC/GEEC pathway. *Cell Host Microbe* **10**, 563–576 (2011).
19. Ghosh, P., Dahms, N. M. & Kornfeld, S. Mannose 6-phosphate receptors: new twists in the tale. *Nature Rev. Mol. Cell Biol.* **4**, 202–213 (2003).
20. Beglova, N. & Blacklow, S. C. The LDL receptor: how acid pulls the trigger. *Trends Biochem. Sci.* **30**, 309–317 (2005).
21. Ohka, S. *et al.* Receptor (CD155)-dependent endocytosis of poliovirus and retrograde axonal transport of the endosome. *J. Virol.* **78**, 7186–7198 (2004).
22. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV serotypes 1–9 mediated gene expression and tropism in mice after systemic injection. *Mol. Ther.* **16**, 1073–1080, (2008).
23. Ran, F. A. *et al.* *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
24. Balazs, A. B. *et al.* Antibody-based protection against HIV infection by vectored immunoprophylaxis. *Nature* **481**, 81–84 (2012).
25. Nonnenmacher, M. & Weber, T. Intracellular transport of recombinant adeno-associated virus vectors. *Gene Ther.* **19**, 649–658 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The authors thank K. Kirkegaard, M. Kay and T. Brummelkamp for critical reading of the manuscript and valuable advice; T. Lerch, J. Tyner, D. Kabat and H. Nakai for assistance with preliminary experiments; H. P. Bächinger for advice and assistance with surface plasmon resonance experiments; Stanford Shared FACS facility and its staff; X. Ji (Stanford Functional Genomics Facility); Stanford mouse facility; T. Doyle for small animal imaging training; L. Popov for guidance in generating immunofluorescent images; G. Fuchs for technical assistance; and members of the Carette and Chapman laboratories for intellectual discussions and support. The work was funded in part by NIH R01 GM066875 (M.S.C.), DP2 AI104557 (J.E.C.) and U19 AI109662 (J.E.C.). J.E.C. is a David and Lucile Packard Foundation fellow.

**Author Contributions** S.P., M.S.C. and J.E.C. were responsible for overall design of the study. S.P. performed the haploid genetic screen, generated isogenic knockout and AAVR complement cell lines, and performed antibody inhibition and AAVR tracking studies. N.L.M. designed soluble AAVR construct and performed binding and soluble AAVR inhibition studies. J.E.C. designed AAVR-generated deletion mutant constructs. A.S.P. performed the wild-type AAV2 infection assay and all *in vivo* studies, under the technical expertise of C.M.N. O.D. was responsible for heterologous overexpression and purification of soluble AAVR. J.D. assisted in the production of *FGFR1*<sup>KO</sup> and *MET*<sup>KO</sup> cell lines. Y.I. performed surface plasmon resonance measurements, L.T.J. generated the *B3GALT6*<sup>KO</sup> cell line, and J.E.W. created the  $\Delta$ C-tail construct. S.P., M.S.C. and J.E.C. wrote the manuscript.

**Author Information** DNA sequencing data have been deposited in the NCBI sequencing read archive under NCBI BioProject PRJNA284536 with BioSample SAMN03703230 (gene-trap control data set) and SAMN04244346 (AAV screen). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.E.C. ([carrette@stanford.edu](mailto:carrette@stanford.edu)) or M.S.C. ([chapmami@ohsu.edu](mailto:chapmami@ohsu.edu)).

## METHODS

**Cell lines and viruses.** All cells were grown in media supplemented with 10% fetal calf serum (FCS) (Sigma), 100 IU ml<sup>-1</sup> penicillin/streptomycin (Sigma) and 2 mM L-glutamine (Sigma), and grown in a humidified incubator at 37°C with 5% CO<sub>2</sub>. HAP1 (ref. 8) cells and K562 cells (American Type Culture Collection (ATCC)) were cultured in complete IMDM media. HT29, U2OS (both obtained from ATCC), Caco-2, A549 (both gifts from L. Popov), HEK-293T (from Thermo Scientific), H1-HeLa (from ATCC), HuH7 (a gift from P. Sarnow), mouse embryonic fibroblasts (a gift from K. Storek) and NIH3T3 (a gift from W. Kaiser) cells were all cultured in complete DMEM media. Raji cells (expressing DC-SIGN) (a gift from E. Harris) were cultured in complete RPMI media. The cell lines have not been authenticated or tested for mycoplasma contamination. All isogenic knockout clones were grown in the same media as parent cell lines. HAP1 cells were used for haploid genetic screens (see later). Purified, titred stocks of AAV serotypes 1, 2, 3B, 5, 6, 8 and 9 were purchased from University of North Carolina Chapel Hill Gene Therapy Center Vector Core. These were all self-complementary AAV vectors encoding a reporter fluorescent gene (either GFP or RFP). Purified titred stocks of AAV9-luciferase were also purchased from this core facility to perform mouse experiments. Adenovirus type 5 vector carrying mCherry (Ad5-RFP) was constructed by cloning mCherry cDNA in the pAd/CMV/V5-DEST gateway vector (Invitrogen) according to the manufacturer's protocol.

**Antibodies.** The following antibodies were used in this study: mouse polyclonal anti-KIAA0319L (ab105385) and rabbit polyclonal anti-giantin (ab24586) (Abcam); rabbit polyclonal anti-TGN46 antibody (NBP1-49643) (Novus Biologicals); mouse monoclonal anti-GAPDH (GT239) (Genetex); rabbit polyclonal anti-FGFR1 (D8E4) and rabbit IgG2a isotype control (Cell Signaling Technology); mouse monoclonal phycoerythrin-conjugated anti-MET antibody (95106) and phycoerythrin-conjugated mouse IgG1 isotype control (R&D Systems, Inc). A high-affinity F-actin, fluorescently labelled probe (Alexa Fluor-660 phalloidin) was used to visualize the cell interior and periphery (Life Technologies).

**Virus infections.** Cells were seeded at 10,000 cells per well (96-well plate) overnight. They were then infected with AAV at a MOI of 20,000 vg per cell (unless otherwise specified) in complete DMEM. Virus infectivity was determined 24 h after infection by measuring transgene expression (RFP, GFP or luciferase) using flow cytometry or bioluminescence. In the case of wild-type AAV2 infection, HeLa wild-type or AAVR<sup>KO</sup> cells were seeded overnight, then infected with wild-type AAV2 (MOI: 1,000 vg per cell) in the presence of wild-type Ad5 (helper virus). RNA was collected using the Ambion Cell-to-CT kit (Thermo Scientific) 24 h after infection, and the generated cDNA was used to perform quantitative reverse-transcriptase PCR (RT-qPCR). mRNA levels of the AAV2-encoded *rep68* gene were measured (as a means to detect viral replication) and normalized to 18S ribosomal RNA. Primers against *rep68* cDNA included: 5'-CCAATTACTTGCTCCCCAAA-3' and 5'-CGTTTACGCTCCGTGAGATT-3'. Primers against 18S rRNA included: 5'-AGAAACGGCTACCACATCCA-3' and 5'-CACCAGACTTGCCCTCCA-3'. Ad5-RFP was used to infect cells to obtain 50–60% transduction (Fig. 4a), and flow cytometry was used to measure RFP expression. All infections were performed in triplicate, and all data presented are representative of at least two independent experiments.

**Haploid genetic screen.** The haploid genetic screen was performed similarly to the protocol described in ref. 8 with minor changes. Briefly, gene-trap virus was used to create a mutagenized HAP1 library. Of this mutagenized library, 100 million cells were infected with AAV2-RFP at a MOI 20,000 of vg per cell. After 48 h, infected cells were sorted by FACS, where RFP-negative cells (approximately 4% of the population) were sorted and grown over a period of 4 days. The resulting sorted cells were then infected again with AAV2 as before, and re-sorted to enrich the RFP-negative (AAV-resistant) population. Thirty-million cells of the resistant population were used for genomic DNA isolation. We performed sequence analysis of gene-trap insertion sites, and the significance of enrichment for each gene in the screen was calculated by comparing how often that gene was mutated and how often the gene carried an insertion in the control data set (owing to random integration). For each gene, a *P* value was calculated using the one-sided Fisher exact test in R. The *P* values were corrected for multiple testing according to the Benjamini and Hochberg method (using the R statistical package), to control for false discovery rate<sup>26</sup>. In the case of KIAA0319L, the *P* value was lower than the software could report. The numerical value was thus set to  $1 \times 10^{-307}$  (smallest non-zero normalized floating-point number R could report).

**Generation of isogenic knockout cell lines.** CRISPR-Cas9 gene editing technology was used to generate isogenic knockout alleles by targeting exonic sequences shared among all protein-coding transcripts of the respective genes as described in ref. 27. The targeted sequences are depicted in Extended Data Table 1, along with the respective mutations. CRISPR sequence targeting oligonucleotides were designed using the Zhang lab CRISPR design tool (<http://crispr.mit.edu>).

Oligonucleotides corresponding to the guide RNA (gRNA) sequences in Extended Data Table 1 were synthesized (Integrated DNA Technologies). gRNA oligonucleotides were directly cloned into Cas9-expressing plasmids pX330 or pX458 (generated by the Zhang lab; obtained from <http://www.addgene.org>; plasmid 42230 or 48138). Respective cells were transiently transfected with gRNA-encoding plasmids (and GFP-expressing pcDNA vector with gRNA-pX330 plasmids) using Fugene (Promega). After 48 h, GFP-expressing cells were subcloned using the BD Influx Cell Sorter at the Stanford Shared FACS facility. They were then expanded over 2 weeks and screened genotypically for the mutated allele by extracting genomic DNA from subclones (using the quick DNA universal 96-kit; Zymo Research), amplifying a 500–700 base-pair (bp) region that encompassed the gRNA-targeted site, and sequencing (ElimBio) the resulting PCR product to identify subclones with knockout mutations. The B3GALT6 isogenic knockout clone was generated using TALENs directed against the nucleotide sequence 5'-TGGCCATGCTGGCCTGGCTG-3', and the reverse complement sequence of 5'-GAGTTCGTGCTCAAGGCGGA-3' in the only exon of B3GALT6 (transcript ENST00000379198) as described previously<sup>28</sup>. One day after transfection, cells were selected with blasticidin S (30 µg ml<sup>-1</sup>, InvivoGen) for 24 h, then stained using anti-heparan sulfate antibody. Cells displaying low staining intensity were subcloned by FACS.

**Construction of plasmids.** To generate the AAVR full-length construct and ΔC-tail, Gibson assembly reaction kit (New England Biolabs) was used to insert the gene of interest into a lentiviral-based vector, pLenti-CMV-Puro-DEST (w118-1) (plasmid 17452), digested with EcoRV to remove the DEST cassette (a gift from E. Campeau)<sup>29</sup>. AAVR and derived AAVR genes were amplified from a KIAA0319L cDNA clone (clone ID 3843301) (GE Dharmacon), but a single nucleotide polymorphism at position 447 was changed from a 'T' to a 'G', allowing the sequence to align to the annotated human genome. The following primers were used to generate PCR products from the human KIAA0319L cDNA to be cloned directly into pLent-CMV-Puro-DEST. AAVR full-length: 5'-ATGTGTGGTGAATTCTGCAGATACCATGGAGAAGAGGCTGGG-3' and 5'-CCGCGCCCACTGTGCTGGATTACTATCGTCGTCATCCTTGTAATCCAGGATCTCTCCCGC-3'; ΔC-tail: 5'-GACTCTAGTCCAGTGTGGTG-3' and 5'-CGGCGCCCACTGTGCTGGATTACTATCGTCGTCATCCTTGTAATCTCCTTTTTCGCTCTTACAAC-3'. Note that reverse primer was designed to incorporate a C-terminal 1 × Flag-tag sequence.

To generate the AAVR deletion constructs, two or three PCR products were generated using the AAVR construct (with Flag-tag) as a template. They were then assembled into the pLenti-CMV-Puro-DEST vector using the Gibson assembly reaction. Primers used to amplify the N-terminal fragments for the following constructs were: ΔMANEC: 5'-GACTCTAGTCCAGTGTGGTG-3' and 5'-CTCATGCGCATCTGTTGAC-3'; ΔPKD1-2: 5'-GACTCTAGTCCAGTGTGGTG-3' and 5'-CAGTTCCTTTATAAAGCTGGGTATGG-3'; ΔPKD2-3: 5'-GACTCTAGTCCAGTGTGGTG-3' and 5'-CTTACGGGGCTCTGGC-3'; ΔPKD3-4: 5'-GACTCTAGTCCAGTGTGGTG-3' and 5'-GTAATCCACAGCTTTGTTTTCAGGTTGCACAAT-3'; miniAAVR: 5'-GACTCTAGTCCAGTGTGGTG-3' and 5'-CTCATGCGCATCTGTTGAC-3'; middle fragment of miniAAVR: 5'-GTCAACAGATGCCAGTGAGGTATCTGCTGGAGAGAGTGTG-3' and 5'-CTTATTGTTTTTCAGGTTGCACAAT-3'.

Primers used to amplify the C-terminal fragments for the following constructs were:

ΔMANEC: 5'-GTCAACAGATGCCAGTGAGACACACTCCTCCAATTCAT-3' and 5'-ATCCAGAGGTTGATTGTGCGAG-3'; ΔPKD1-2: 5'-CCATACCCAGTTATAAAGGAAGTCCCCCTGTGGCCAACG-3' and 5'-ATCCAGAGGTTGATTGTGCGAG-3'; ΔPKD2-3: 5'-GCCAGAGCCCCGTAAAGCTCCTCAGGCAGATGC-3' and 5'-ATCCAGAGGTTGATTGTGCGAG-3'; ΔPKD3-4: 5'-GTGAACAAAGCTGTGGATTACCCACCTATAGCCAAAGATACTG-3' and 5'-ATCCAGAGGTTGATTGTGCGAG-3'; ΔPKD4-5: 5'-ATTGTGCAACCTGAAAACAATAAGAAGCTGGTGGAGATCATCTTGGATATC-3' and 5'-ATCCAGAGGTTGATTGTGCGAG-3'; miniAAVR: 5'-ATTGTGCAACCTGAAAACAATAAGTGTGAGTGGAGCGTGTATATG-3' and 5'-ATCCAGAGGTTGATTGTGCGAG-3'.

AAVR PKD domains 1–5 (residues 311–787) were expressed in *E. coli* using the pMAL expression system (New England Biolabs). A bacmid, created from a pFastBac dual vector containing the cDNA for the KIAA0319L ectodomain fused to a C-terminal influenza haemagglutinin (HA)-tag was a gift from M. van Oers, and obtained with the assistance of M. Waye<sup>30</sup>. cDNA coding for PKD domains 1–5 was cloned out of the pFastBacDual expression vector and inserted into the pMAL-c5X vector, using 5'-GTATCTGCTGGAGAGAGTGTCCAGATAACC-3' and 5'-CAGGTTGTTTTCTCGCAGGTCACCTGGGATCAGGTTTCAC-3', then expressed in NEBexpress cells (New England Biolabs). This resulted in a fusion protein comprised of a mannose-binding protein (MBP) tag and AAVR



PKD domains 1–5 (referred to as: soluble AAVR). MBP was specifically used as an affinity tag for ease of purification.

To create AAVR fusion constructs, Ci-MPR-tail, LDLR-tail and PVR-tail, the Gibson assembly reaction was used to fuse amplified miniAAVR without its C-terminal to the C-terminal of the respective proteins, and insert it into the pLenti-CMV-Puro-DEST vector. Primers used for amplification and insertion included: miniAAVR without C-terminal and transmembrane domain for Ci-MPR-tail: 5'-GACTCTAGTCCAGTGTGGTG-3' and 5'-CTTATTGTTTTCAGGTTGCACAAT-3'; MPR C-terminal and transmembrane: 5'-ATTGTGCAACCTGAAAACAATAAGGCTGTGGGAGCTGTGC-3' and 5'-CGGCCGCCACTGTGC-3'; miniAAVR without C-terminal and transmembrane domain for LDLR-tail or PVR-tail: 5'-GACTCTAGTCCAGTGTGGTG-3' and 5'-CTTATTGTTTTCAGGTTGCACAAT-3'; LDLR or PVR C-terminal and transmembrane: 5'-ATTGTGCAACCTGAAAACAATAAG-3' and 5'-TAAATCCAGCACAGTGGCGGCCG-3'.

**Generation of stable cell lines.** Lentiviral transduction was used to create stable cell lines expressing a selected gene of interest under a CMV promoter. Using Gibson assembly reaction, the respective genes of interest (see 'construction of plasmids' section) were inserted into the pLenti-CMV-Puro-DEST vector, and used as described previously<sup>29</sup>. Lentivirus was produced using HEK293 cells and used to transduce the respective cell lines overnight. Cells stably expressing the gene of interest were selected by treatment with 1–3  $\mu\text{g ml}^{-1}$  puromycin over 2 days (InvivoGen). A lentivirus carrying the mCherry (RFP) gene was used as a control for AAVR complementation in AAVR<sup>KO</sup> cells.

**Flow cytometry.** All flow cytometry was performed at the Stanford Shared FACS facility. To perform the haploid genetic screen, FACS was carried out on a FACS Aria flow cytometer (BD). To measure virus transgene expression (RFP/GFP) in all other experiments, cells were trypsinized 24 h after infection and a LSRII-UV flow cytometer (BD) was used to detect fluorescent cells. For cell surface staining, cells were trypsinized and washed using FACS buffer (PBS supplemented with 2% FCS, 1 mM EDTA and 0.1% sodium azide). They were subsequently incubated for 40 min at 4 °C with the respective primary antibodies at a 1:50 dilution (see 'Antibodies' section), washed, and incubated for a further 40 min at 4 °C with Alexa488- or Alexa594-conjugated secondary antibodies (1:500 dilution; if the primary was not conjugated) (Life Technologies). This was followed by a final wash and resuspension of cells in FACS buffer before reading fluorescence. All data presented are representative of at least two independent experiments. Data were analysed and assembled using FlowJo software (TreeStar Inc).

**Immunoblot analysis.** Cell pellets of  $2 \times 10^6$  cells were lysed with Laemmli SDS sample buffer containing 5%  $\beta$ -mercaptoethanol and boiled for 10 min at 96 °C. Lysates were separated by SDS-PAGE using the Mini-Protean system (Bio-Rad) on 4–15% polyacrylamide gradient gels (Bio-Rad). Proteins were transferred onto polyvinylidene fluoride membranes (Bio-Rad) using the Bio-Rad Transblot protein transfer system in a semi-wet preparation. Membranes were blocked by incubating with PBS containing 5% non-fat milk for 1 h at room temperature. Membranes were subsequently incubated overnight at 4 °C with primary antibodies at a dilution of 1:1000 (anti-KIAA0319L antibody) or 1:2,000 (anti-GAPDH antibody) in blocking buffer. Membranes were washed 3 times for 5 min using wash buffer (PBS with 0.1% Tween-20), and further incubated in horseradish peroxidase (HRP)-conjugated secondary antibodies (anti-mouse and anti-rabbit 1:5,000 in blocking buffer) (GeneTex) for 1 h at room temperature. After another set of three washes, antibody-bound proteins were visualized on film using the West Pico and Extended Duration chemiluminescence peroxide solutions (Thermo Scientific).

**Immunofluorescence.** Cells were seeded overnight at 40,000 cells per well onto LabTekII glass chamber slides (Thermo Scientific). They were washed once with PBS, and either treated or fixed immediately with 4% paraformaldehyde for 15 min. They were washed three times with PBS before being incubated for 1 h at room temperature with primary antibodies against the respective proteins at a dilution of 1:100 (anti-KIAA0319L and anti-TGN46) or 1:200 (anti-giantin) in immunofluorescence blocking buffer (PBS with 3% BSA, 1% saponin and 1% Triton X-100). Cells were then washed three times in PBS, and incubated for a further hour in DAPI stain (1:500) and fluorescently tagged secondary antibodies (Alexa488 anti-mouse and Alexa594 anti-rabbit; Life Technologies) at a dilution of 1:300. Cells were washed a final three times in PBS, and 5  $\mu\text{l}$  of Vectashield (Vector Laboratories Inc) was applied to each slide chamber before a glass cover slip (VWR International) was placed over slide to mount samples. Cells were visualized directly with a Zeiss LSM 700 confocal microscope.

**ELISA.** Purification of the soluble AAVR was achieved through amylose-based MBP affinity chromatography (GE Healthcare). ELISA plates (Corning Costar) were coated overnight at 4 °C with 50  $\mu\text{l}$  AAV2 virus-like particles at 2.5  $\mu\text{g ml}^{-1}$  in 100 mM  $\text{NaHCO}_3$  (pH 9.6). Plates were then washed with TBST buffer (0.05% Tween-20 in TBS) and blocked with 3% BSA in TBST for 1 h at room temperature. Subsequent washing was followed by incubation with soluble AAVR or MBP

control at the indicated concentrations for 2 h at room temperature. Anti-MBP-HRP (1:500, 1 h incubation at room temperature) was used to detect soluble AAVR and MBP controls, requiring no secondary antibody. Samples were developed with 1-Step Ultra TMB-ELISA substrate as per the manufacturer's instructions (Thermo Scientific) and optical density assayed by microplate reader (Molecular Devices SpectraMax M2<sup>2</sup>) at 450 nm. Curve fitting was performed in SigmaPlot v12.5 (Systat Software, Inc). All data presented are representative of at least three independent experiments.

**Surface plasmon resonance analysis of binding.** Surface plasmon resonance analysis was carried out using a Biacore X instrument (GE Healthcare) and a flow rate of 10  $\mu\text{l min}^{-1}$  at 20 °C in HBS-P buffer (10 mM HEPES (pH 7.5), 150 mM NaCl and 0.005% surfactant P20). His-tagged soluble AAVR (His-tagged MBP fusion with AAVR PKD domains 1–5) at various concentrations was mixed with His-tagged MBP to a total concentration of 0.2  $\mu\text{M}$  in 10 mM sodium acetate buffer (pH 4.0) and immobilized on a CM5 sensor chip by amide coupling. MBP at 0.2  $\mu\text{M}$  was sufficient to block nonspecific binding to the dextran. For the analysis of binding affinity, all curves were measured in triplicate and were fitted with a Langmuir 1:1 binding model (BIAevaluation software, GE Healthcare).

**Antibody inhibition assay.** Wild-type HeLa cells were seeded in 96-well plates at 10,000 cells per well overnight. Anti-AAVR antibody (ab105385) or IgG isotype control (both from Abcam) were incubated with cells (at concentrations ranging from 0.5 to 50  $\mu\text{g ml}^{-1}$  in DMEM media) for 1 h at 4 °C. Cells were then infected with AAV2-luciferase at a MOI of 1,000 vg per cell, and left for 24 h at 37 °C. A luciferase assay kit (E1500, Promega) was used to detect bioluminescence, with measurements being taken on the Promega GloMax luminometer. Notably, the storage buffers of both antibodies did not contain preservatives such as azide that could interfere with the assay. All data presented are representative of two independent experiments.

**Competitive inhibition assay.** HeLa cells were seeded in 96-well plates at 10,000 cells per well overnight. Purified soluble AAVR or MBP control was then introduced to the medium at the specified concentrations. Cells were transduced with AAV2-GFP at a MOI of 7,500 vg per cell and incubated for 24 h at 37 °C. This was followed by trypsinization and measuring transgene expression by flow cytometry. For immunofluorescence imaging, the concentration of soluble AAVR and MBP controls was 0.1  $\mu\text{M}$ , and transduction was performed using 7,000 vg per cell. At 24 h post-transduction, cells were incubated with 1  $\mu\text{g ml}^{-1}$  Hoechst stain (Thermo Scientific) in PBS for 10 min at 37 °C, before washing with PBS and subsequent fluorescent imaging (Nikon Eclipse Ti-E). All data presented are representative of two independent experiments.

**Tracking surface-bound AAVR using anti-AAVR antibodies.** These experiments were performed similarly to Ci-MPR tracking assays, as described in ref. 31. AAVR<sup>KO</sup> cells with or without overexpression of AAVR or  $\Delta\text{C}$ -tail were incubated at 4 °C with anti-AAVR antibodies (approximately 25  $\mu\text{g ml}^{-1}$ ) for 1 h. Cells were then washed three times with PBS and transferred to 37 °C for specific time points (2, 10, 30 and 60 min), at which time they were fixed with 4% paraformaldehyde for 15 min. Following fixation, immunofluorescence staining (as described earlier) was performed to visualize AAVR endocytosis. All data presented are representative of two independent experiments.

**Ethics statement and animal studies.** All the experiments involving animals were conducted in strict accordance with the Institutional Animal Care and Use Committee of Stanford University. Mice were housed in a Stanford University vivarium, accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International. Mice were housed in irradiated disposable caging (Innovive Inc) with bi-weekly cage changes. Mice were provided with irradiated food and ultraviolet-irradiated acidified water. Health surveillance was performed via trimester testing of dirty bedding CD1 sentinels (Charles River Laboratories). Sentinels were consistently negative for mouse parvovirus, minute virus of mice, mouse hepatitis virus, rotavirus, mouse encephalomyelitis virus, Sendai virus, mouse adenovirus 1 and 2, ectromelia virus, lymphocytic choriomeningitis virus, pneumonia virus of mice, reovirus 3, *Mycoplasma pulmonis*, and endo- and ectoparasites. No statistical methods were used to predetermine sample size. In our animal study protocol, we state that the number of animals in each experimental group varies, and is based on similar previous study<sup>32</sup>. Randomization was not used to allocate animals to experimental groups and the investigators were not blinded to allocation during experiments and outcome assessment.

**AAV infection in mice.** TALEN technology was used to create AAVR isogenic knockout FVB mice (purchased from Cyagen Biosciences). TALEN-targeted sequences were 5'-TGGGAGTCAAGCCAAGTC-3' and 5'-GCCAGGATATTGTTGGCAGA-3'. Two founder males were mated to FVB/NCrl (Charles River Laboratories) females. After three rounds of breeding, wild-type (*Aavr*<sup>+/+</sup>), heterozygous (*Aavr*<sup>+/-</sup>) and homozygous AAVR<sup>KO</sup> (*Aavr*<sup>-/-</sup>) mice were generated, determined by genotyping. All genotypes (wild-type, heterozygous

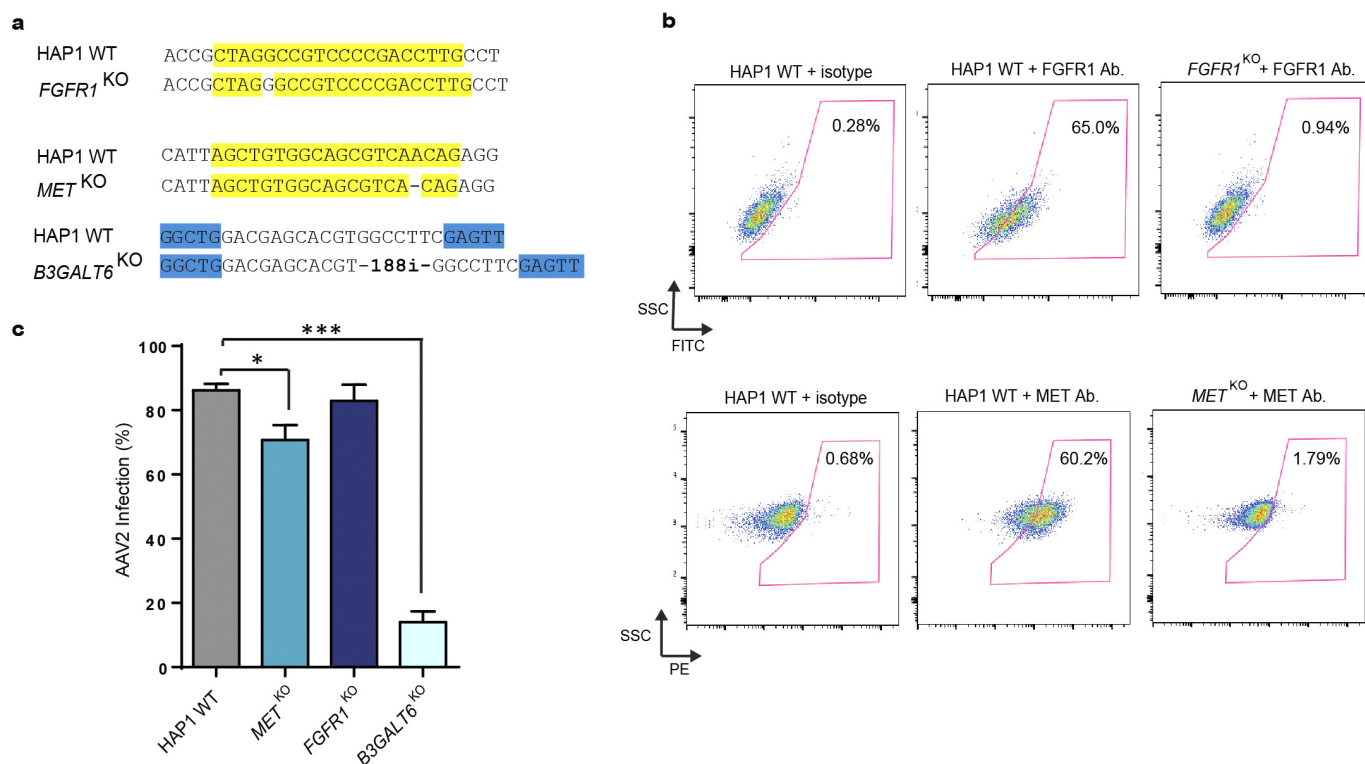


and knockout) were obtained in the expected Mendelian ratios after breeding. At 5 weeks of age, 10 female and 9 male animals were used to examine the effect of *Aavr* KO on AAV infection *in vivo*. Animals from each group (*Aavr*<sup>+/+</sup>, *n* = 7 (2 litter mates and 5 purchased FVB mice); *Aavr*<sup>+/-</sup>, *n* = 4; *Aavr*<sup>-/-</sup>, *n* = 4 and uninfected mice, *n* = 4) were injected intraperitoneally with  $1 \times 10^{11}$  viral genomes of AAV9-luciferase in 200  $\mu$ l of PBS. All of the mice recovered from the injection quickly without loss of mobility or interruption of grooming activity. *Aavr*<sup>+/+</sup> and *Aavr*<sup>-/-</sup> mice were found to be significantly different in two independent experiments.

***In vivo* bioluminescence imaging.** The mice were anaesthetized with 2% isoflurane and oxygen. The D-luciferin substrate (Biotium) was injected intraperitoneally (3.3  $\mu$ g per mouse). After 10 min, the mice were then placed in a light-tight chamber, and images were generated using a cryogenically cooled charge-coupling device camera IVIS 100 (Xenogen), recording bioluminescence at 1, 10, 60 and 100 s. The visual output represents the average radiance as the number of photons emitted per second per cm<sup>2</sup> as a false colour image where the maximum is red and the minimum is dark blue. All animals were imaged on a schedule of 3, 7, 10 and 14 days after AAV vector injection. At each time-point a 'region of interest' was designated surrounding each animal in order to quantify the radiance (photons s<sup>-1</sup> cm<sup>-2</sup> sr<sup>-1</sup>) being released by luciferase activity. This region was kept the same for each mouse and at each time point. The mean and standard deviation of radiance measurements were determined for each mouse group at each time point.

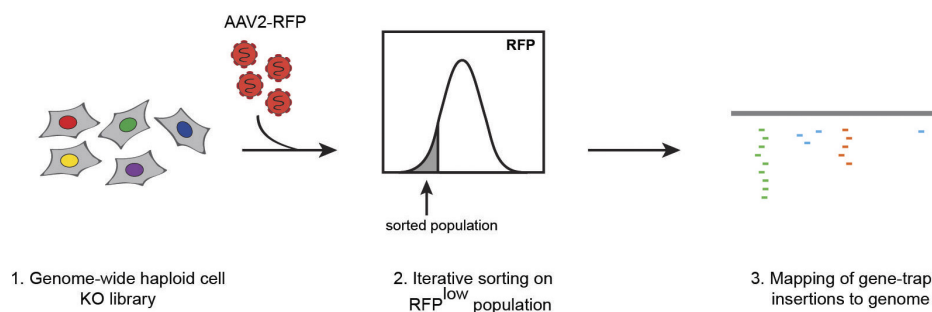
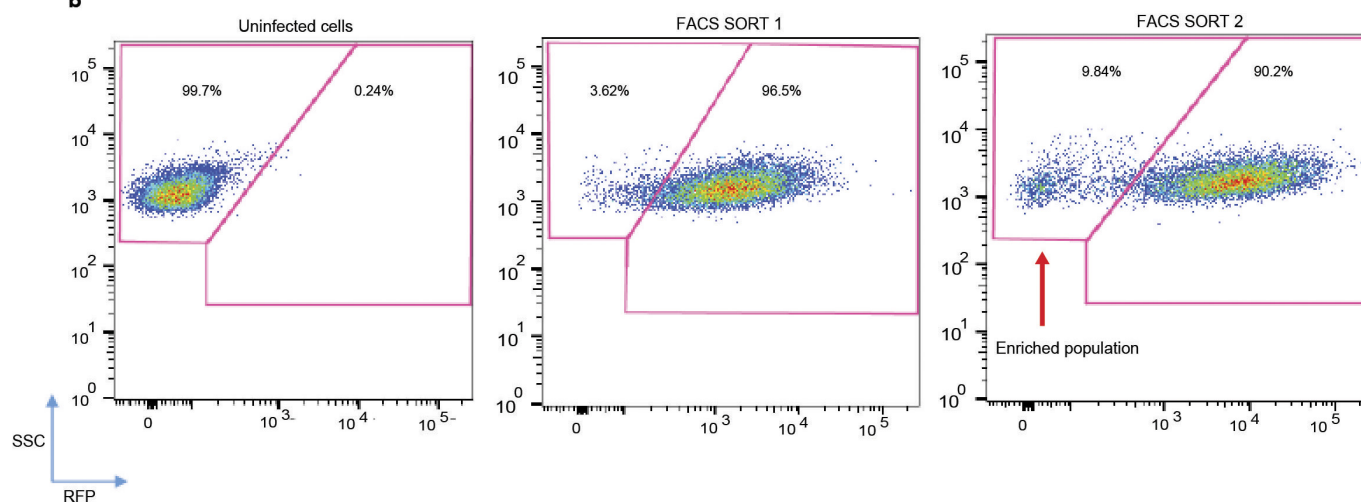
**Statistics.** The unpaired parametric two-sided Student's *t*-test was used for statistical calculations involving two group comparisons in all tissue-culture-based experiments (\**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001), with a Welch post-correction accounting for different standard deviations. An unpaired two-sided Mann-Whitney *t*-test was used for statistical calculations involving two group comparisons in *in vivo* experiments. GraphPad Prism was used for statistical calculations.

26. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. A Stat. Soc.* **57**, 289–300 (1995).
27. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**, 2281–2308 (2013).
28. Sanjana, N. E. *et al.* A transcription activator-like effector toolbox for genome engineering. *Nature Protocols* **7**, 171–192 (2012).
29. Campeau, E. *et al.* A versatile viral system for expression and depletion of proteins in mammalian cells. *PLoS ONE* **4**, e6529 (2009).
30. Holster, S. *et al.* Expression of the dyslexia candidate gene Kiaa0319-like in insect cells. *J. Biochem. Mol. Biol. Post Gen. Era* **2**, 45–52 (2013).
31. Seaman, M. N. Cargo-selective endosomal sorting for retrieval to the Golgi requires retromer. *J. Cell Biol.* **165**, 111–122 (2004).
32. Jae, L. T. *et al.* Virus entry. Lassa virus entry requires a trigger-induced receptor switch. *Science* **344**, 1506–1510 (2014).



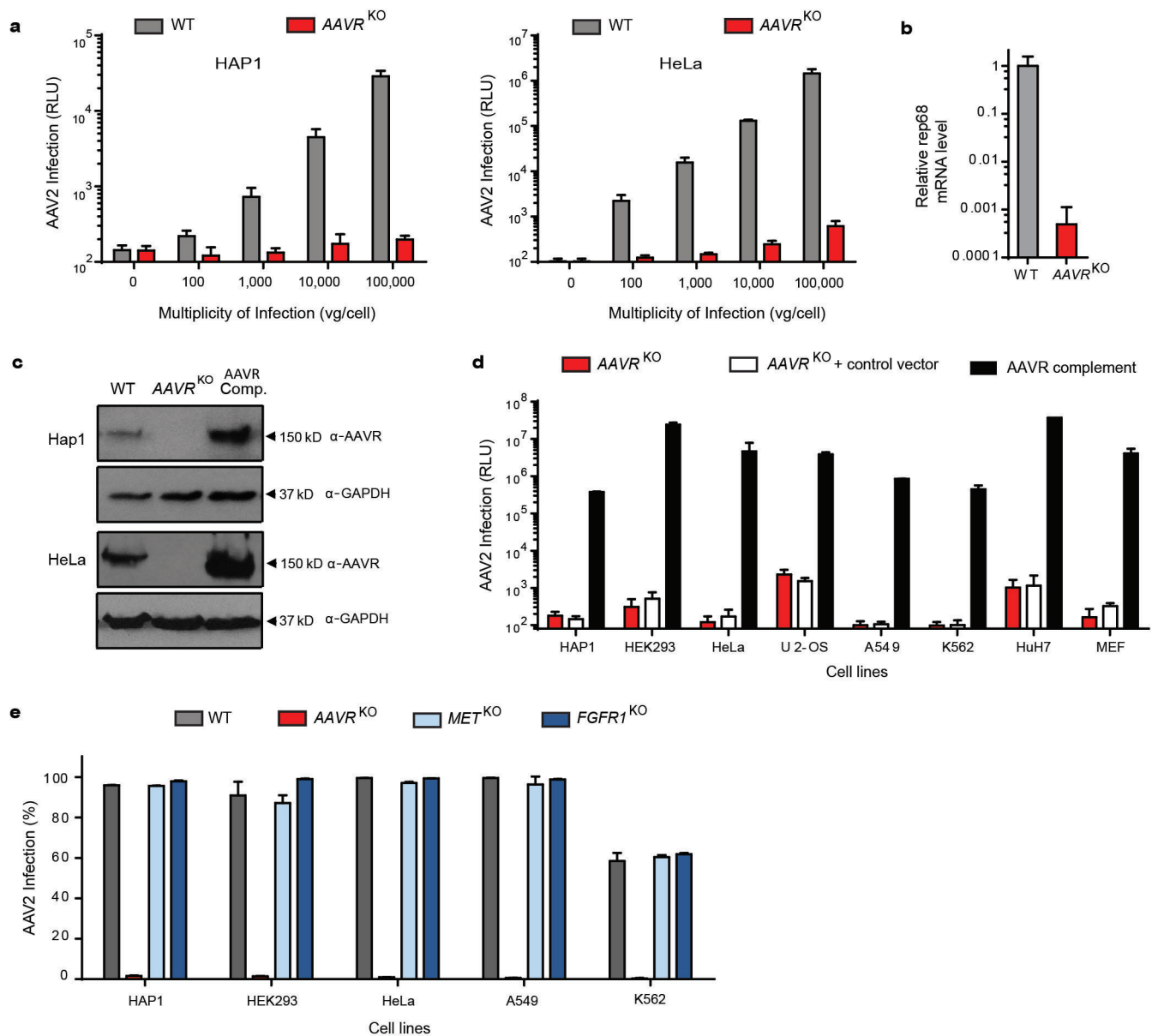
**Extended Data Figure 1 | Surface molecules *FGFR1* and *MET* are not essential for AAV2 infection.** **a**, Region of *FGFR1*, *MET*, or *B3GALT6* genes (previously identified co-receptors/attachment factors<sup>5-7</sup>) targeted by CRISPR gRNA or TALENs in wild-type HAP1 cells, and the resulting genotypes of derived knockout cell lines (see Extended Data Table 1 for full sequences). All CRISPR- or TALEN-created mutations disrupt the open reading frame of the targeted gene. **b**, Surface staining for the respective receptors in respective cell lines. Isotype antibodies for the

receptor antibodies (Ab) were used as controls. **c**, AAV2-RFP infection (MOI 5,000 vg per cell; measured after 24 h) of wild-type and knockout cell lines. Data depict the mean and s.d. for triplicate infections. \* $P < 0.05$ , \*\*\* $P < 0.001$ ; analysed using unpaired parametric two-sided Student's *t*-test, with Welch post-correction. SSC, side scatter. Fluorescently labelled antibody conjugates, fluorescein isothiocyanate (FITC) and phycoerythrin (PE), were used to visualize surface receptors.

**a****b**

**Extended Data Figure 2 | Haploid unbiased genetic screen evaluating host factors important for AAV2 infection.** **a**, A schematic depicting the strategy for the AAV2 genetic screen. A library of mutagenized haploid HAP1 cells was created with a retroviral gene-trap vector, and subsequently infected with AAV2-RFP (MOI: 20,000 vg per cell) for 24 h. RFP-negative cells were sorted using FACS to isolate cells with mutations

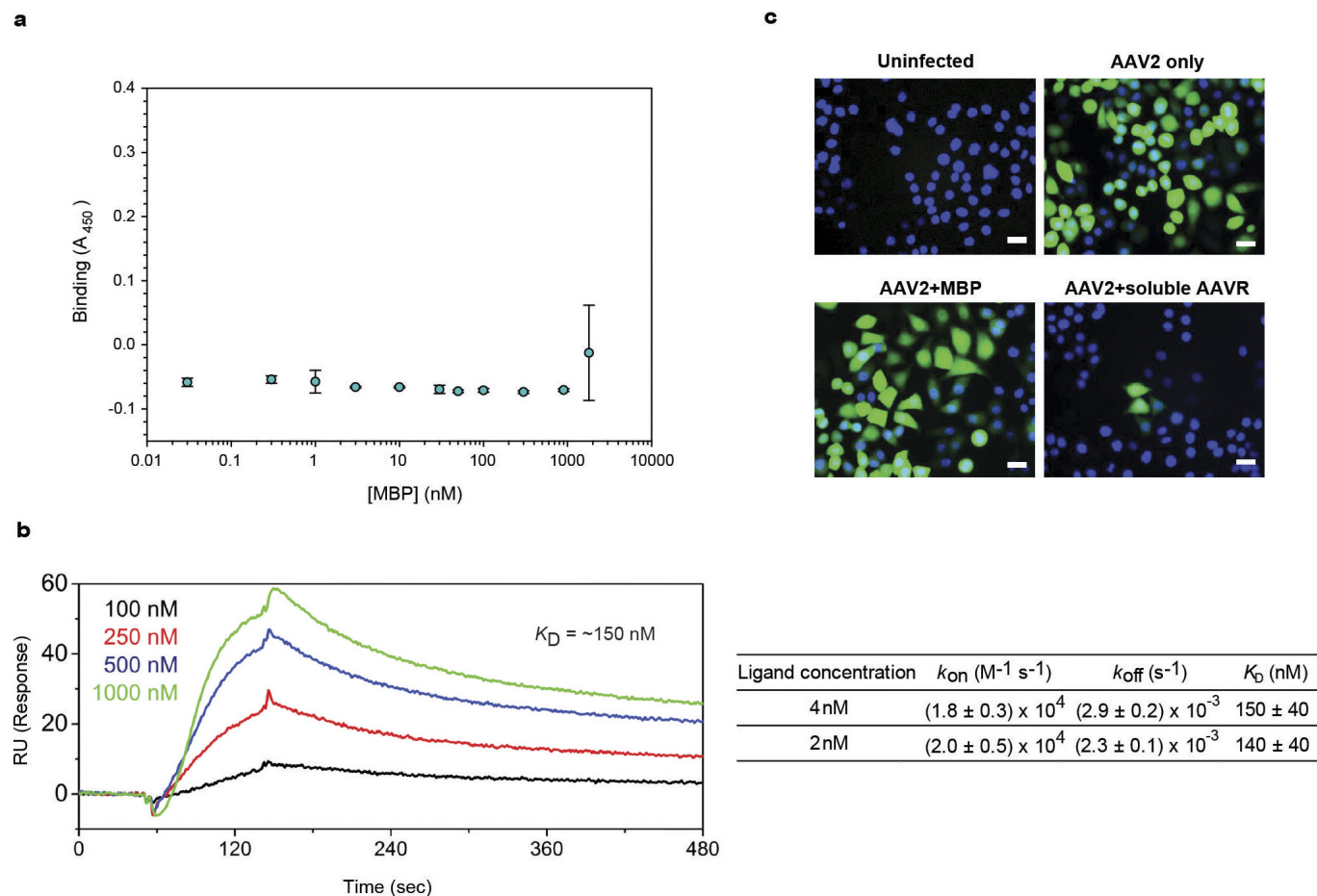
in genes essential for AAV2 infection. These cells were re-infected for a second iteration of selection. DNA was then extracted from this enriched population and sequenced to map specifically where the gene-trap insertions occurred that resulted in the mutation. **b**, The gating strategy for the FACS-based AAV2 screen.



**Extended Data Figure 3 | AAVR is a critical host factor for AAV2 infection.** **a**, Effect of AAVR isogenic knockout (AAVR<sup>KO</sup>) upon AAV2-luciferase infection, evaluated in HAP1 and HeLa cell background from a MOI of 100 to 100,000 vg per cell. **b**, RT-qPCR to detect wild-type AAV2 infection in wild-type HeLa or AAVR<sup>KO</sup> cells. Cells were infected with wild-type AAV2 and adenovirus (helper virus required for AAV2 replication), and AAV2 *Rep68* mRNA levels were measured to assess AAV2 infection. **c**, Immunoblot analysis evaluating AAVR expression in wild-type, AAVR<sup>KO</sup> and AAVR<sup>KO</sup> overexpressing AAVR (AAVR Comp.) cell

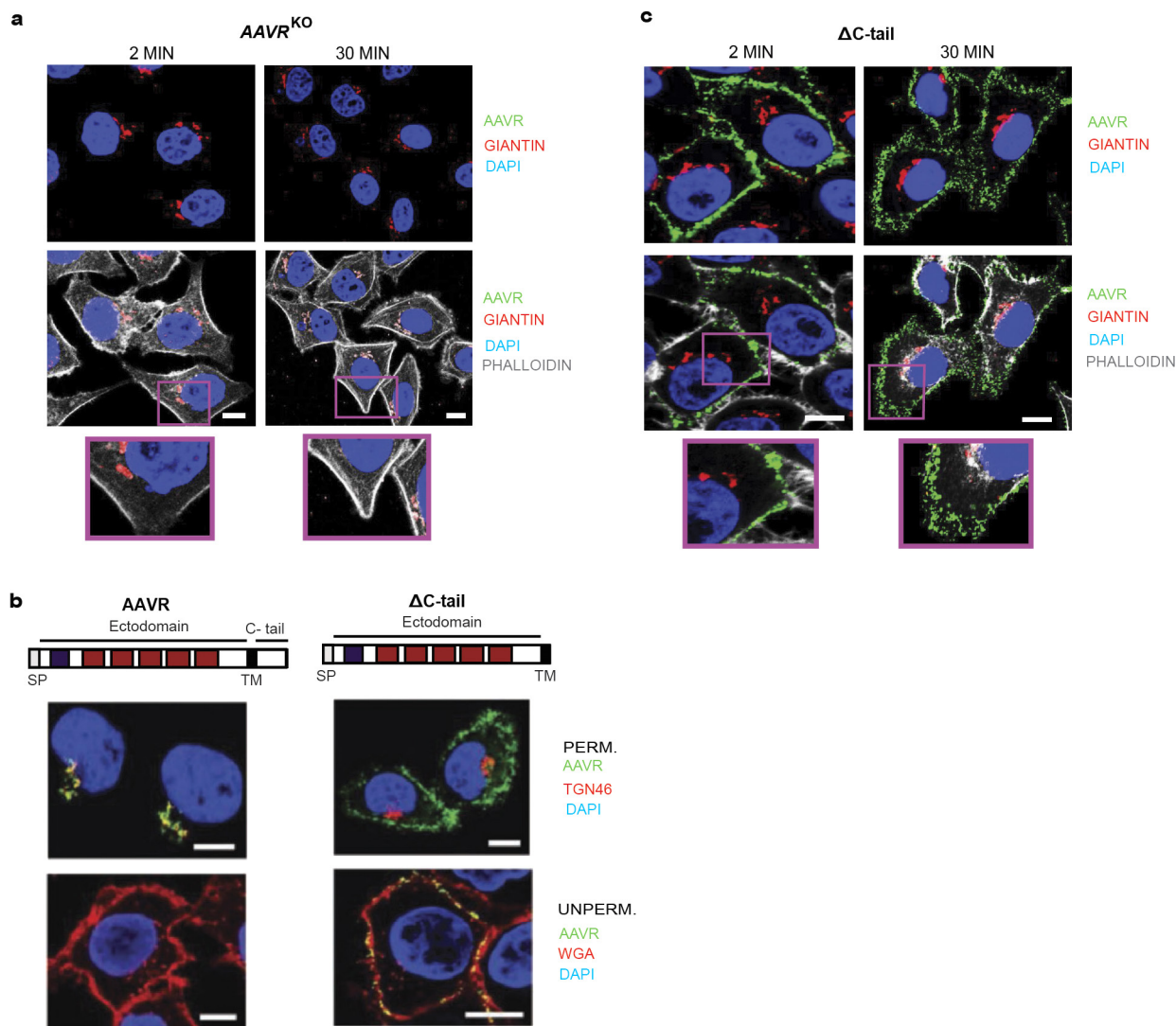
lines of HAP1 and HeLa origin. GAPDH was immunoblotted as a control. AAVR (predicted 115 kilodaltons (kDa)) appears at 150 kDa owing to six glycosylation sites. **d**, AAV2-luciferase infection (MOI 20,000 vg per cell; measured after 24 h) in AAVR<sup>KO</sup> cells stably complemented with AAVR or control lentiviral vector, evaluated in several AAV2-susceptible human and mouse cell lines. **e**, Comparison of AAV2-RFP infection (MOI: 20,000 vg per cell; measured after 24 h) in wild-type, AAVR<sup>KO</sup>, MET<sup>KO</sup> and FGFR1<sup>KO</sup> cells, evaluated in several AAV2-susceptible human cell lines. Data depict the mean and s.d. for triplicate infections.





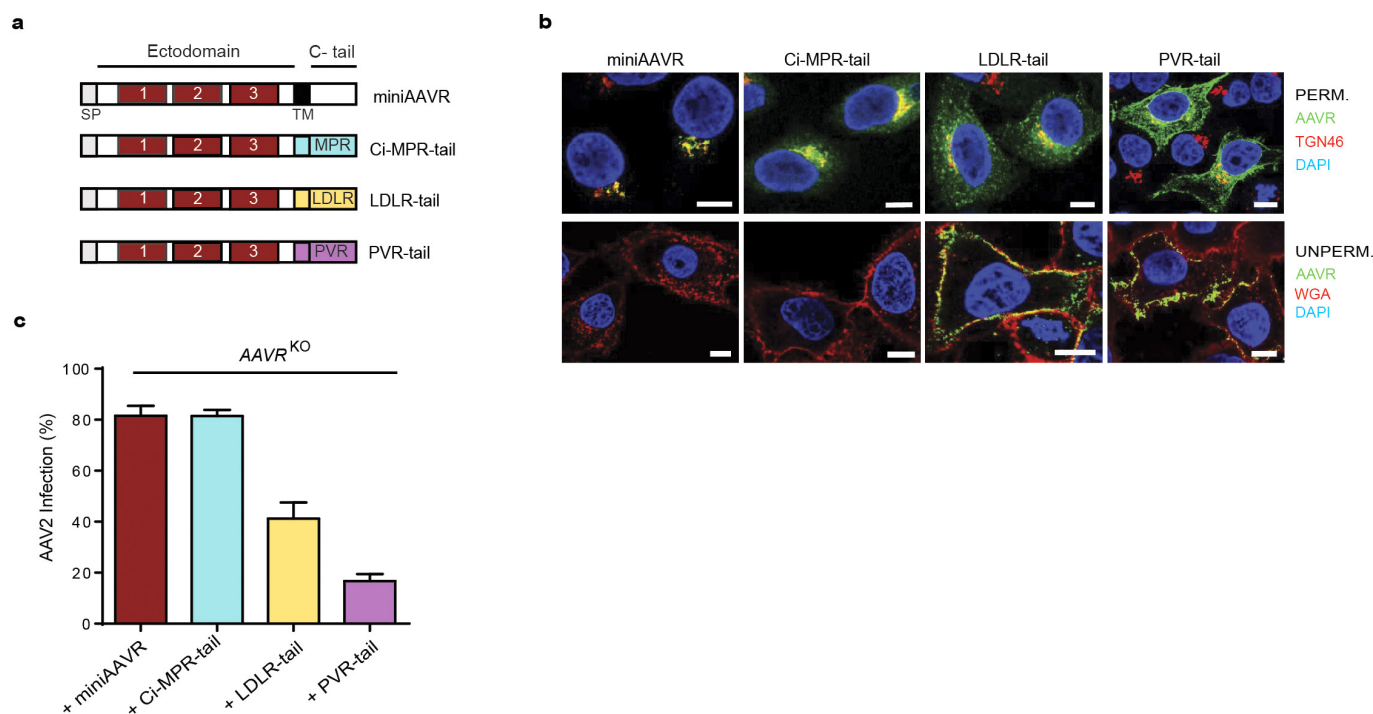
**Extended Data Figure 4 | AAVR specifically binds to AAV2.** **a**, ELISA measurement of the binding to AAV2 particles of MBP at concentrations of 0.05–2,000 nM. This serves as a control to the ELISA data depicted in Fig. 2c. **b**, Representative surface plasmon resonance sensograms (collected in triplicate), with a ligand (AAVR) concentration of 4 nM and an analyte (AAV2) concentration as indicated, to measure binding of

AAV2 particles to AAVR. **c**, Simultaneous addition to cells of AAV2–GFP particles with soluble AAVR or MBP (both at 0.1  $\mu$ M) to evaluate binding effect of AAVR on AAV2 infection. Fluorescence was imaged 24 h after infection. These data complement Fig. 2d. Data in **a** depict the mean and s.d. for triplicate infections. Scale bars, 50  $\mu$ m. RU, response units.



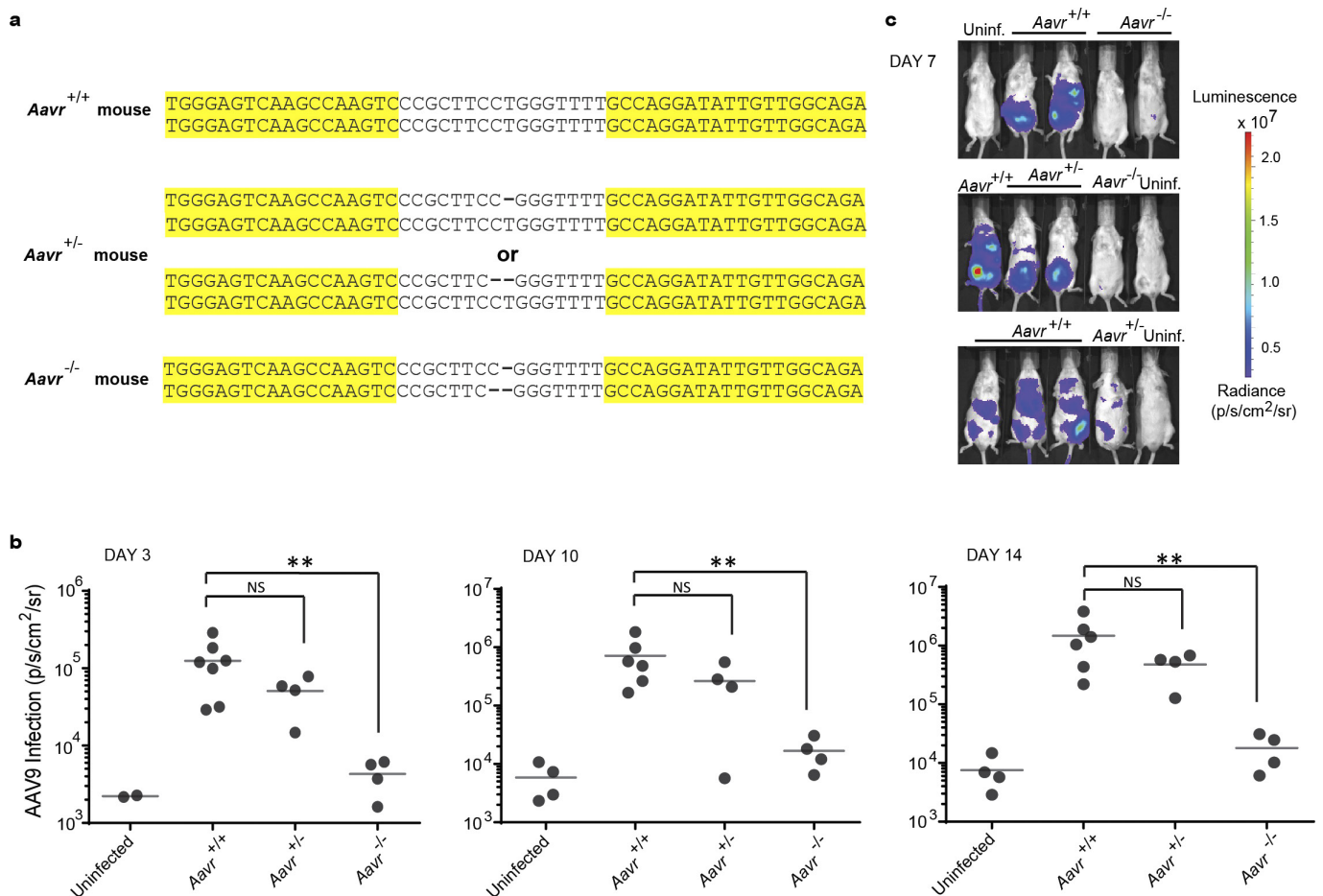
**Extended Data Figure 5 | AAVR  $\Delta$ C-tail is detected at the cell surface and does not endocytose to the TGN.** AAVR<sup>KO</sup> cells (a) or  $\Delta$ C-tail-expressing cells (c) were incubated with anti-AAVR antibodies for 1 h at 4°C, washed and then transferred to 37°C. At respective time points, cells were fixed and

antibody-bound AAVR was visualized. These data complement Fig. 3b. **b**, Permeabilized and unpermeabilized immunostaining of full-length AAVR and  $\Delta$ C-tail when expressed in AAVR<sup>KO</sup> cells. These data complements Fig. 3c. Scale bars, 10  $\mu$ m.



**Extended Data Figure 6 | AAVR endocytosis is crucial for AAV2 infection.** **a**, Schematic of the miniAAVR and domain-swapped derivatives probing the localization of AAVR through the swapping of the AAVR C-tail with that of well-characterized recycling receptors: Ci-MPR (traffics from plasma membrane through endosomes to the TGN), LDLR and PVR (both traffic from plasma membrane to endosomal compartments but

are not reported to traffic to TGN). **b**, Corresponding permeabilized and unpermeabilized immunofluorescence images of constructs depicted in **a** when expressed in AAVR<sup>KO</sup> cells. **c**, AAV2-RFP infection (MOI: 20,000 vg per cell; measured after 24 h) in AAVR<sup>KO</sup> cells stably expressing constructs depicted in **a**. Data depict the mean and s.d. for triplicate infections. Scale bars, 10 μm.



# Extended Data Figure 7 | AAVR is essential for AAV infection *in vivo*.

**a**, Genotypes of FVB mice littermates used to perform *in vivo* studies. AAVR knockout mice (*Aavr*<sup>-/-</sup>) were bred from heterozygous (*Aavr*<sup>+/-</sup>) parent mice; *Aavr*<sup>+/-</sup> and *Aavr*<sup>-/-</sup> mice display frameshift mutations in targeted genes in 1 and 2 alleles, respectively. Sequences recognized by the TALENs are displayed in yellow. **b**, AAV9-luciferase infection for all

infected mice depicted for day 3, 10 and 14 (day 7 is shown in Fig. 4d).

**c**, Bioluminescence in all wild-type (*Aavr*<sup>+/+</sup>), *Aavr*<sup>+/-</sup> and *Aavr*<sup>-/-</sup> FVB mice 7 days after AAV9-luciferase infection (does not include those shown in Fig. 4b). Radiance range of  $2 \times 10^5$  to  $1 \times 10^7$  photons s<sup>-1</sup> cm<sup>-2</sup> sr<sup>-1</sup>. Significance was determined using unpaired two-sided Mann-Whitney *t*-test; \*\**P* < 0.01.



Extended Data Table 1 | Indel mutations in genes of isogenic knockouts created by CRISPR-Cas9 or TALEN genome engineering

Gene	CRISPR/TALEN region (bold)	Cell line	Indel mutation in allele
<i>KIAA0319L</i> (AAVR)	CCAG <b>GTGACGTAGTTACACCTATAGTG</b>	HEK293	CC----- <b>AGTTACACCTATAGTGACAC</b> CCAG----- <b>CTATAGTGACAC</b> CCAG <b>TG</b> -- <b>G</b> TAGTTACACCTATAGTGACAC
		A549	CCAG <b>T</b> -- <b>ACGTAGTTACACCTATAGTGACAC</b> CCAG <b>TGAACGTAGTTACACCTATAGTGACAC</b>
		U2OS	C----- <b>CCTATAGTGACA</b> CCAG <b>TG</b> ----- <b>173i</b> ----- <b>ACGTAGTTAC</b>
		HuH7	CCAG <b>TG</b> ----- <b>271d</b> ----- CCAG <b>T</b> -- <b>ACGTAGTTACACCTATAGTGACAC</b> CCAG <b>TG</b> -- <b>G</b> TAGTTACACCTATAGTGACAC
		K562	CC----- <b>AGTTACACCTATAGTGACAC</b> CCAG <b>TGAC</b> ----- <b>69d</b> -----
	CTT <b>GCTTTTGCTTCAGCGTTCTGTGG</b>	HAP1	CTT <b>GCTTTTGCTTCAGCGTTCTGTGGTTG</b>
		HeLa	CTT <b>GCTTTTGCTT</b> ----- <b>147d</b> ----- CTT <b>GCTTTTGCTTCAGCG</b> -- <b>TCTGTGGTTG</b>
<i>Au040320</i> (mouse AAVR)	GACT <b>CTGCCTGCCACGCTCTATGGTGG</b>	MEF	GACT <b>CTGCCTGCCACGCTCTA</b> -- <b>GGTGGCTG</b> GACT <b>CTGCCTGCCACGCT</b> ---- <b>GGTGGCTG</b>
<i>FGFR1</i>	CACCG <b>CTAGGCCGTCCCCGACCTTGCC</b>	HAP1	CACCG <b>CTAGGCCGTCCCCGACCTTGCCCTG</b>
		HEK293	CACC----- <b>CGACCTTGCCCTG</b>
		A549	CACCG <b>CTAGGCCGTCCCCGACCTTGCCCTG</b>
		K562	CACCG <b>CTAG</b> -- <b>CCGTCCCCGACCTTGCCCTG</b>
		HeLa	CACCG <b>CTA</b> ----- <b>CCTG</b>
<i>MET</i>	TCATT <b>AGCTGTGGCAGCGTCAACAGAG</b>	HAP1	TCATT <b>AGCTGTGGCAGCGTCA</b> -- <b>CAGAGGGA</b>
		HEK293	TCATT <b>AGCTGTGGCAGCGTCA</b> <b>AA</b> CAGAGGGA
		A549	TCATT <b>AGCTGTGGCA</b> -- <b>27d &amp; 5i</b> -- <b>CCT</b>
		K562	TCATT <b>AGCTGTGGCAGCGTCA</b> -- <b>CAGAGGGA</b> TCATT <b>AGCTGTGGCAGCGTCA</b> <b>AA</b> CAGAGGGA
		HeLa	TCATT <b>AGCTGTGGCAGCGTCA</b> -- <b>AGAGGGA</b> TCATT <b>AGCTGTGGCAGCGTCA</b> <b>AA</b> CAGAGGGA
<i>B3GALT6</i>	<b>TGGCCATGCTGGCCTGGCTG</b> GACGAGCAGTGCC TTC <b>GAGTTCGTGCTCAAGCGGA</b>	HAP1	<b>GCTG</b> GACGAGCAGT-- <b>188i</b> -- <b>GGCCTTCGAG</b>

d, deletion; i, insertion.

# Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells

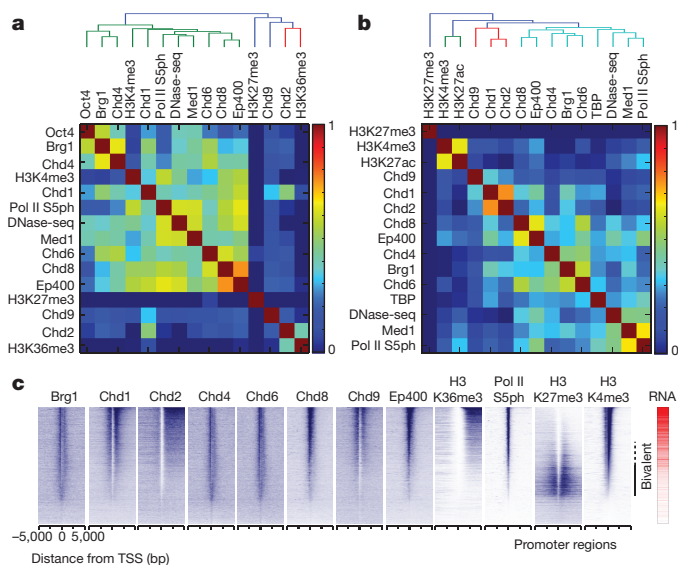
Maud de Dieuleveult<sup>1\*</sup>, Kuangyu Yen<sup>2,3\*</sup>, Isabelle Hmitou<sup>1\*</sup>, Arnaud Depaux<sup>1\*</sup>, Fayçal Boussouar<sup>1†</sup>, Daria Bou Dargham<sup>1</sup>, Sylvie Jounier<sup>1</sup>, Hélène Humbertclaude<sup>1</sup>, Florence Ribierre<sup>4</sup>, Céline Baulard<sup>4</sup>, Nina P. Farrell<sup>3</sup>, Bongsoo Park<sup>3</sup>, Céline Keime<sup>5</sup>, Lucie Carrière<sup>1</sup>, Soizick Berlivet<sup>1</sup>, Marta Gut<sup>6</sup>, Ivo Gut<sup>6</sup>, Michel Werner<sup>1</sup>, Jean-François Deleuze<sup>4</sup>, Robert Olaso<sup>4</sup>, Jean-Christophe Aude<sup>1</sup>, Sophie Chantalat<sup>4</sup>, B. Franklin Pugh<sup>3§</sup> & Matthieu Gérard<sup>1§</sup>

ATP-dependent chromatin remodellers allow access to DNA for transcription factors and the general transcription machinery, but whether mammalian chromatin remodellers<sup>1–3</sup> target specific nucleosomes to regulate transcription is unclear. Here we present genome-wide remodeller–nucleosome interaction profiles for the chromatin remodellers Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Brg1 and Ep400 in mouse embryonic stem (ES) cells. These remodellers bind one or both full nucleosomes that flank micrococcal nuclease (MNase)-defined nucleosome-free promoter regions (NFRs), where they separate divergent transcription. Surprisingly, large CpG-rich NFRs that extend downstream of annotated transcriptional start sites are nevertheless bound by non-nucleosomal or subnucleosomal histone variants (H3.3 and H2A.Z) and marked by H3K4me3 and H3K27ac modifications. RNA polymerase II therefore navigates hundreds of base pairs of altered chromatin in the sense direction before encountering an MNase-resistant nucleosome at the 3' end of the NFR. Transcriptome analysis after remodeller depletion reveals reciprocal mechanisms of transcriptional regulation by remodellers. Whereas at active genes individual remodellers have either positive or negative roles via altering nucleosome stability, at polycomb-enriched bivalent genes the same remodellers act in an opposite manner. These findings indicate that remodellers target specific nucleosomes at the edge of NFRs, where they regulate ES cell transcriptional programs.

We applied a genome-wide remodeller–nucleosome interaction assay<sup>4</sup> (MNase digestion to define nucleosomes, followed by remodeller chromatin immunoprecipitation and sequencing (ChIP-seq)) to ES cells, focusing on the 5' ends of genes (Extended Data Fig. 1 and Supplementary Table 1). We first examined remodeller co-enrichment with other factors such as RNA polymerase (Pol) II, selected histone marks and transcription factors, over broad (500-base-pair (bp)) windows centred on DNase-I hypersensitive sites (DHSs) (that is, promoters and enhancers;  $n = 138,582$ ) (Fig. 1a). High Pearson correlation scores were observed among the remodellers Brg1 (also known as Smarca4), Ep400, Chd1, Chd4, Chd6 and Chd8, suggesting that these factors tend to occupy the same genomic regions in ES cells. When we focused on active promoter regions within DHSs, most remodellers were correlated with components of the general transcription machinery, including Pol II S5ph and TBP (Fig. 1b and Extended Data Fig. 2).

We next examined remodeller distribution in more detail by focusing on annotated transcriptional start sites (TSSs) (Fig. 1c and Extended Data Fig. 3). Remarkably, some remodellers such as Brg1, Chd4 and

Chd6 bound similar nucleosome positions at all active genes, regardless of their H3K4me3 enrichment (which is a mark of transcriptional activity), whereas others, such as Chd1, Chd2, Chd9 and Ep400, were tightly linked to histone 3 Lys4 trimethylation (H3K4me3)/transcription levels. Chd8 had an intermediate pattern. Chd1 and Chd2, which are both related to *Saccharomyces cerevisiae* (yeast) Chd1, showed markedly different distributions. Whereas Chd1 is present near the 5' ends of genes, the Chd2-nucleosome enrichment pattern encompassed the entire transcription unit and shared high correlation with H3K36me3 (Fig. 1a, c and Extended Data Fig. 2). This is consistent with how yeast Chd1 works<sup>5,6</sup>, and thus mammalian Chd2 and yeast Chd1 may be functionally equivalent.

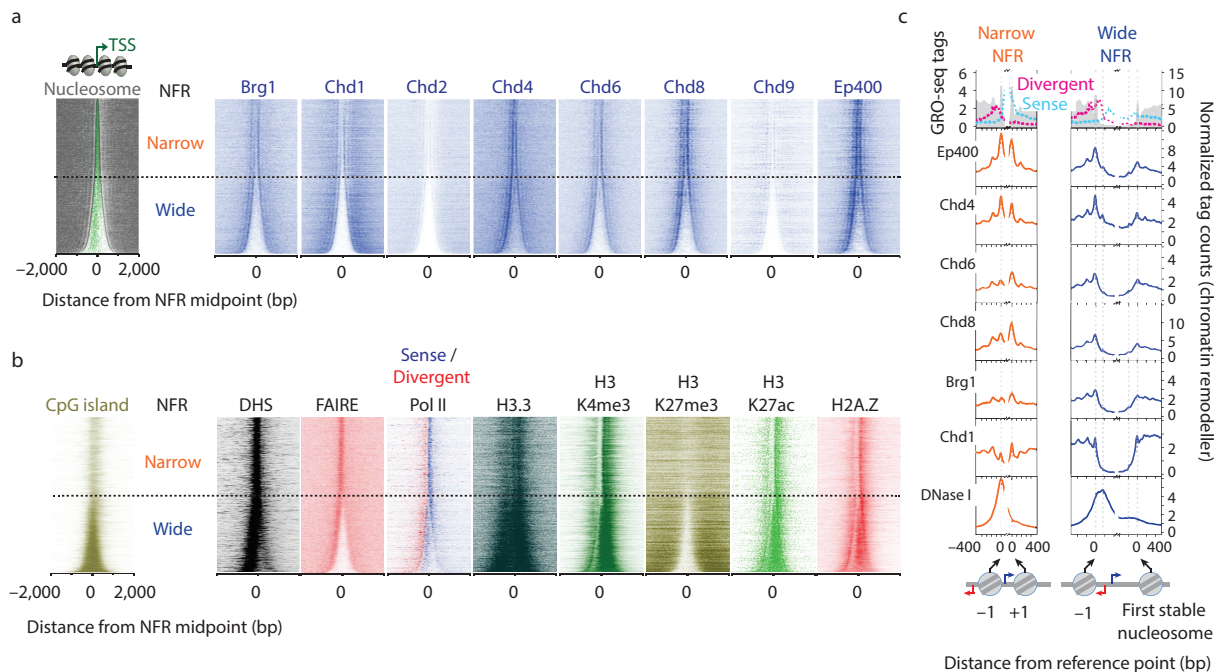


**Figure 1 | Correlated occupancies across remodeller-bound nucleosomal regions.** **a**, Heat map representing Pearson correlations between remodellers and other factors within 500 bp of 138,582 DHS midpoints. **b**, Same as in **a** but for 16,300 promoter-like, H3K4me3-, TBP- and Pol II S5ph-positive DHSs. **c**, Distribution of remodeller–nucleosome interactions (MNase ChIP-seq tags for the indicated remodellers in blue) aligned at 14,623 individual RefSeq TSSs (rows), sorted by H3K4me3 levels. Corresponding RNA expression levels (red) are shown.

<sup>1</sup>Institute for Integrative Biology of the Cell (I2BC), IBITECS, CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette, France. <sup>2</sup>Key Laboratory of Zebrafish Modeling and Drug Screening for Human Diseases of Guangdong Higher Education Institutes, Department of Developmental Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China. <sup>3</sup>Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>4</sup>Centre National de Génotypage, Institut de Génétique, CEA, 91057 Evry, France. <sup>5</sup>IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), INSERM, U964, CNRS, UMR7104, Université de Strasbourg, 67404 Illkirch, France. <sup>6</sup>Centre Nacional D'Anàlisi Genòmica, 08028 Barcelona, Spain. <sup>†</sup>Present address: INSERM, U823; Université Grenoble Alpes; Institut Albert Bonniot Grenoble, 38700 Grenoble, France.

\*These authors contributed equally to this work.

§These authors jointly supervised this work.



**Figure 2 | Patterns of remodeler–nucleosome interactions and chromatin features around promoter NFRs.** **a**, Distribution of remodeler–nucleosome interactions, as in Fig. 1c, except aligned by NFR midpoint and sorted by NFR length. Standard MNase-defined nucleosomes (grey) and TSS (green) are shown. Narrow and wide NFRs are delineated by the dashed line. **b**, Same as in **a** for other genomic features. **c**, Averaged distribution of remodeler–nucleosome interactions

We next investigated more closely the relationship between individual remodeler-bound nucleosomes and all nucleosomes defined by MNase-resistant mononucleosome-sized DNA fragments<sup>7</sup>. Plots of individual genes were aligned by their NFR midpoint and sorted by NFR width into narrow and wide groups (Fig. 2a). We validated the experimental approach and its improved resolution by comparison to an existing sonication-based (rather than MNase) ChIP-seq approach<sup>8</sup> (Extended Data Fig. 4). Importantly, this sonication-based method, which reports on both nucleosomal and non-nucleosomal interactions, demonstrated that Chd4 was not bound within NFRs in a non-nucleosomal manner.

At narrow NFRs, Ep400 and Chd4 crosslinked predominantly nucleosomes  $-1$  and  $+1$  that flank the NFR (Fig. 2a). Chd6, Chd8 and Brg1 interacted predominantly with  $+1$  nucleosomes, and at lower levels with  $-1$  and  $-2$ . Chd1 was also enriched at  $+1$ , and had a diffuse distribution on several additional nucleosomes on both sides of the TSS (Fig. 2a, c). Thus, at short NFRs, the first nucleosome ( $+1$ ) encountered by Pol II after release from the pause state is one that is highly enriched with remodelers. These remodelers might have a role in the passage of Pol II through these nucleosome barriers.

At wide NFRs, Ep400 and Chd4 were preferentially bound to  $-1$  nucleosomes (Fig. 2a), and relatively less to the first detectable full nucleosome downstream of the NFR. More notably, Chd6, Chd8 and Brg1 had shifted from their preferential binding to  $+1$  at narrow NFRs towards a predominant enrichment at the  $-1$  position of wide NFRs. These NFRs also define the boundaries of CpG islands, as reported previously<sup>9</sup>. Thus, mammalian remodelers interact with nucleosomes in a position-specific manner, with a distribution pattern adapting to the local chromatin architecture and DNA composition.

The pattern of remodeler-bound nucleosomes contrasted with TSSs, including divergent Pol II transcripts defined by global run-on sequencing (GRO-seq)<sup>10</sup>, which generally stayed towards the upstream side of NFRs (that is, distal to annotated gene bodies) (Fig. 2b). DHS and FAIRE<sup>11</sup> patterns, which demarcate chromatin accessibility,

from **a** and **b** at narrow and wide NFRs, aligned to the dyad of  $-1$  (left portion of each graph) or the first MNase-resistant nucleosome downstream of the noncanonical chromatin (right portion). Standard nucleosomes (grey fill) and GRO-seq RNA (blue and red dashed lines) are shown. A gap in the NFR midpoint was introduced to account for variations in NFR length inside each class.

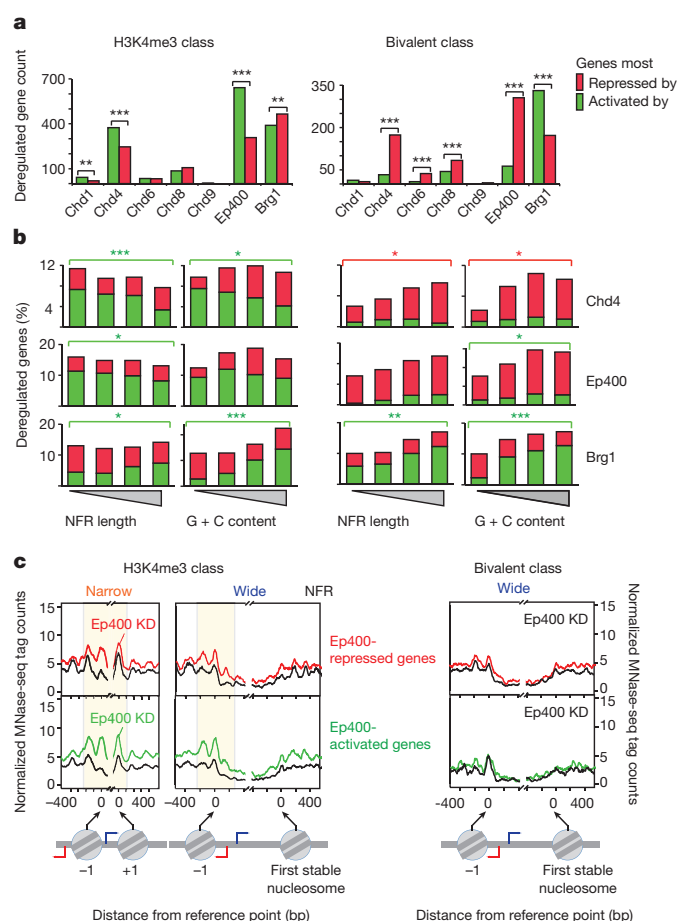
matched the narrower regions of annotated TSSs, rather than reflecting the dimensions of MNase-defined NFRs. Thus, DNase I (or FAIRE) and MNase define regions of differing dimensions. At the enzyme concentrations used and DNA fragment sizes analysed, DNase I (and FAIRE) released histone-free regions (termed HFRs) generating a positive signal, whereas MNase destroyed HFRs and noncanonical chromatin, thereby generating a lower signal. We find that promoter HFRs have a roughly fixed width ( $<115$  bp). Where NFRs are narrow (Fig. 2a), HFRs and NFRs are essentially the same. At wide NFRs, HFRs are embedded in the upstream portion of NFRs that are variably wider, CpG-enriched, and contain remarkably noncanonical chromatin (being DNase I resistant but MNase sensitive).

We examined the distribution of histone variants and marks in NFRs, measured by standard ChIP-seq<sup>12–16</sup>. Narrow NFRs had H3.3, H2A.Z, H3K4me3 and H3K27ac enriched primarily at the bordering  $+1$  nucleosome, with H3K4me3 extending to nucleosome  $+3$ . Some enrichment occurred upstream, largely commensurate with the level of divergent transcription (Fig. 2b). Remarkably, at wide NFRs, these variants and marks were largely restricted to the noncanonical chromatin region downstream of promoter HFRs, but still within NFRs.

A subset of less active genes is marked by a combination of H3K4me3 and H3K27me3, defining them as bivalent<sup>17</sup> (Fig. 1c). Notably, bivalency does not predominate on the same nucleosomes (that is, H3K4me3 and H3K27ac are enriched in NFRs, whereas H3K27me3 resides downstream, over full nucleosomes) (Fig. 2b). Thus, wide mammalian NFRs are largely chromatinized with non-nucleosomal (MNase-defined) transcription-associated histone modifications and variants that may be spatially adjacent to repressive chromatin.

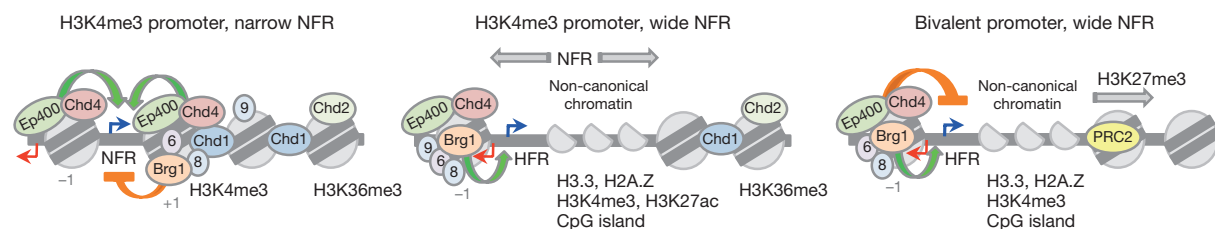
At narrow NFRs, the Ep400-bound and Chd4-bound  $-1$  nucleosomes separated sense-directed Pol II from upstream divergently transcribed Pol II (Fig. 2c, top left). This  $-1$  nucleosome also represented the peak of DNase I hypersensitivity (bottom left). These remodelers therefore might be involved in structural reorganization, ejection or repositioning of  $-1$  to regulate sense and divergent transcription





**Figure 3 | Remodellers differentially regulate active versus bivalent genes.** **a**, The number of genes in which RNA was either downregulated (green) or upregulated (red) after remodeller depletion by  $>1.5$ -fold is shown for H3K4me3-only or bivalent genes.  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$  (two-sample test for equality of proportions with continuity correction). **b**, The percentages of H3K4me3 (left) or bivalent (right) genes upregulated (red) or downregulated (green) by remodeller depletion are shown in four subgroups based on NFR length, as defined in Fig. 2, and in quartiles for C + G content. Statistical significance metrics (described in **a**) are colour-matched and applied to the first and last group. **c**, Averaged nucleosome distribution (MNase-seq) after control (black) or Ep400 (colour) knockdown at H3K4me3-only (narrow versus wide NFR) and bivalent gene groups. Top and bottom panels represent genes that are upregulated (red) or downregulated (green) ( $>1.5$ -fold) after Ep400 knockdown (KD).

(examined below). At bivalent genes, Brg1, Chd4, Chd6, Chd8 (weakly) and Ep400 (weakly) bound specific nucleosomes at a level commensurate with their transcription levels, akin to what was observed at non-bivalent 'H3K4me3-only' gene class (Extended Data Fig. 5). By contrast, Chd1 was enriched only at the H3K4me3 class.



**Figure 4 | Model of how remodellers might regulate distinct classes of genes in ES cells.** Three gene classes are indicated, having remodeller-bound nucleosomes (coloured circles on top of grey circles) at specific positions relative to the TSS (horizontal blue arrow). MNase-sensitive

To investigate how remodellers, bound at these distinct classes of genes, are involved in transcription regulation in ES cells, we depleted each remodeller using short hairpin RNA (shRNA) vectors (Extended Data Fig. 6) and profiled messenger RNA expression or used a published deletion data set<sup>18</sup> for Brg1 (Fig. 3a). We observed that Chd4, Ep400 and Brg1, among the tested remodellers, were the most required for transcriptional expression, in both the H3K4me3 and bivalent classes. Ep400 and Chd4 were predominantly involved in transcriptional activation of H3K4me3 promoters, whereas Brg1 showed a preference for repression. At bivalent promoters, Chd4, Chd6, Chd8 and Ep400 were mostly involved in transcriptional repression, whereas Brg1 counteracted transcriptional repression, as previously described<sup>18</sup>. Loss-of-function of the other remodellers resulted in more limited changes in gene expression. These results were validated by reverse transcriptase quantitative PCR (RT-qPCR) using two different shRNA vectors for each remodeller (Extended Data Fig. 7).

Since CpG islands help determine remodeller requirements<sup>19</sup>, we counted the percentage of genes most regulated by Chd4, Ep400 and Brg1 as a function of C + G content or NFR width, which is related to C + G (and CpG) content (Fig. 3b). Chd4 and Ep400 preferentially activated H3K4me3 genes with narrow NFRs (low CpG content), whereas at bivalent promoters the trend was doubly reversed: genes were preferentially repressed at wide NFRs (high CpG content). Brg1 preferentially activated genes with long NFRs, and repressed genes with narrow NFRs (low CpG content), at both H3K4me3 and bivalent classes. Thus, remodellers and chromatin architecture commonly have reciprocal relationships between H3K4me3 and bivalent genes.

To understand how specific remodellers might regulate genes, we used ATAC-seq (assay for transposase-accessible chromatin-sequencing)<sup>20</sup> to examine regional chromatin accessibility after remodeller depletion. Brg1 depletion resulted in decreased ATAC-seq signal at bivalent promoters (Extended Data Fig. 8), whereas Chd4 depletion produced the opposite effect. Thus, Brg1 maintains accessible chromatin and Chd4 restricts it. Despite Ep400 positively regulating H3K4me3 promoters and negatively regulating bivalent promoters, we detected no effect on regional accessibility by Ep400 using ATAC-seq, indicating that it acts by another mechanism.

To further examine how Ep400 might work, we turned to MNase-seq after Ep400 depletion. Remarkably, its depletion resulted in increased MNase-resistance particularly at the  $-1$  nucleosome, where Ep400 was enriched at both positively and negatively regulated genes (Fig. 3c). This effect was most evident at H3K4me3 promoters, which are bound by high Ep400 levels, compared to bivalent promoters (Fig. 3c and Extended Data Fig. 3). Ep400 therefore may act to alter the structure of the  $-1$  nucleosome.

We also examined the consequences of remodeller depletion on Pol II occupancy at promoters, using ChIP-exo. We found that depletion of either Ep400 or Chd4 resulted in a reduction of Pol II levels at the H3K4me3 promoters they activate (Extended Data Fig. 9), showing that these remodellers contribute to Pol II recruitment at subsets of active promoters.

Three contrasting stereotypes of remodeller control of gene expression in mouse ES cells arise from the data (Fig. 4), although not all

noncanonical chromatin, having histone variants and active marks, is shown as half circles. Curved green and orange ribbons indicate transcriptional activation and repression, respectively. Single digit numbers denote corresponding Chd remodellers.



genes fall into these stereotypes. First, active (H3K4me3-only) genes with narrow NFRs are flanked by nucleosomes bound and destabilized by positive-acting Chd4 and Ep400. The +1 nucleosome is further engaged with negatively acting Brg1. Other remodellers bind there as well, but their function is less clear. Further downstream, Chd2-nucleosome interactions may be using the H3K36me3 mark to organize nucleosomes analogously to Chd1 or Isw1b in budding yeast<sup>6</sup>.

The second stereotype has similar but not identical remodeller-nucleosome interactions as the first, such that Brg1 acts positively through -1 instead of negatively through +1. This stereotype also has wide CpG-rich NFRs that are chromatinized with remodelled non-canonical chromatin or partial nucleosomes (for example, hexasomes, tetrasomes or half-nucleosomes), and includes short upstream HFRs where bidirectional transcription originates. The third stereotype is similar to the second but is enriched with bivalent genes having H3K4me3 within the NFR, and H3K27me3 and polycomb downstream within genic nucleosomal arrays. Thus bivalency is spatially separated on the same gene. Two trends emerge: an activating remodeller in one class of genes is an inhibitor remodeller in the other class; and within the same class, an activating remodeller can be counteracted by an inhibitor remodeller. Taken together, remodellers work together at specific nucleosome positions adjacent to promoter region NFRs to elicit proper gene control.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 6 January; accepted 1 December 2015.**

**Published online 27 January 2016.**

- Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**, 490–503 (2013).
- Becker, P. B. & Workman, J. L. Nucleosome remodeling and epigenetics. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).
- Cairns, B. R. The logic of chromatin architecture and remodelling at promoters. *Nature* **461**, 193–198 (2009).
- Yen, K., Vinayachandran, V., Batta, K., Koerber, R. T. & Pugh, B. F. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* **149**, 1461–1473 (2012).
- Simic, R. *et al.* Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *EMBO J.* **22**, 1846–1856 (2003).
- Smolle, M. *et al.* Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nature Struct. Mol. Biol.* **19**, 884–892 (2012).
- Teif, V. B. *et al.* Genome-wide nucleosome positioning during embryonic stem cell development. *Nature Struct. Mol. Biol.* **19**, 1185–1192 (2012).
- Whyte, W. A. *et al.* Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221–225 (2012).
- Fenouil, R. *et al.* CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.* **22**, 2399–2408 (2012).
- Min, I. M. *et al.* Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev.* **25**, 742–754 (2011).
- Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* **17**, 877–885 (2007).
- Yildirim, O. *et al.* A system for genome-wide histone variant dynamics in ES cells reveals dynamic MacroH2A2 replacement at promoters. *PLoS Genet.* **10**, e1004515 (2014).
- Yukawa, M. *et al.* Genome-wide analysis of the chromatin composition of histone H2A and H3 variants in mouse embryonic stem cells. *PLoS ONE* **9**, e92689 (2014).
- Hu, G. *et al.* H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* **12**, 180–192 (2013).
- Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).
- Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
- Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
- Ho, L. *et al.* esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nature Cell Biol.* **13**, 903–913 (2011).
- Ramirez-Carrozzi, V. R. *et al.* A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114–128 (2009).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10**, 1213–1218 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. C. Andrau, S. Ravens, L. Tora and M. de Chaldée for critical reading, J. B. Charbonnier's team for sharing material, A. Krebs, T. Ye, I. Davidson, R. Guerois and F. Ochsenbein for discussions, and A. Martel for computational assistance. This work was supported by the CEA, and by grants from the Association pour la Recherche sur le Cancer (grant 3164), the Agence Nationale de la Recherche (ANR-05-BLAN-0396), the Fondation pour la Recherche Médicale (FRM), the National Institutes of Health (grant HG004160, to B.F.P.) and from Southern Medical University (grant B1000465, to K.Y.).

**Author Contributions** M.D.D., I.H., A.D., F.B., D.B.D., S.J. and H.H. built cell lines and performed ChIP-seq and shRNA transfections. I.H., C.B. and R.O. conducted transcriptome experiments and analysis. F.R. and S.C. performed FAIRE-seq, ATAC-seq and RT-qPCR experiments. M.Ge. and S.C. carried out MNase-seq experiments. S.B. provided the pHyper shRNA vector. K.Y., M.Ge., L.C., C.K., B.P. and J.-C.A. conducted bioinformatics analysis. N.P.F. performed ChIP-exo and DNA sequencing at PSU. I.G., M.Gu. and J.-F.D. supervised DNA sequencing at CNAG and CNG. M.Ge., K.Y., S.C., M.W. and B.F.P. supervised the project. M.Ge., K.Y. and B.F.P. co-wrote the manuscript.

**Author Information** Data have been deposited in Gene Expression Omnibus (GEO) under accession numbers GSE64825 (ChIP-seq, ATAC-seq, MNase-seq and ChIP-exo data) and GSE64819 (microarray). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Ge. ([matthieu.gerard@cea.fr](mailto:matthieu.gerard@cea.fr)), K.Y. ([kuangyuyen@smu.edu.cn](mailto:kuangyuyen@smu.edu.cn)) or B.F.P. ([bfp2@psu.edu](mailto:bfp2@psu.edu)).

## METHODS

No statistical methods were used to predetermine sample size. Experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

**Knock-in of a TAP-tag in the genes encoding the remodellers through homologous recombination in ES cells.** The recombineering technique<sup>21</sup> was adapted to construct all targeting vectors for homologous recombination in ES cells. Retrieval vectors were obtained by combining 5' miniarm (NotI/SpeI), 3' miniarm (SpeI/BamHI) and the plasmid PL253 (NotI/BamHI). SW102 cells<sup>21</sup> containing a BAC encompassing the carboxy-terminal part of the gene encoding the remodeller, were electroporated with the SpeI-linearized retrieval vector. This allowed the subcloning of genomic fragments of approximately 10 kilobases (kb) comprising the last exon of the gene encoding each remodeller. The next step was the insertion of a TAP-tag into the subcloned DNA, immediately 3' to the coding sequence. The TAP-tag was (Flag)<sub>3</sub>-TEV-HA for Chd1, Chd2, Chd4, Chd6, Chd8, Ep400, Brg1 and 6His-Flag-HA for Chd9. We first inserted the TAP-tag and an AscI site into the PL452 vector, to clone 5' homology arms as SalI/AscI fragments into the PL452TAP-tag vector. 46C ES cells were electroporated with NotI-linearized targeting constructs and selected with G418. In all cases, G418-positive clones were screened by Southern blot. Details on the Southern genotyping strategy, as well as sequences of primers and plasmids used in this study are available on request. Correctly targeted ES cell clones were karyotyped, and the expression of each tagged remodeller was controlled by western blot analysis, using antibodies against Flag and haemagglutinin (HA) epitopes (see Extended Data Fig. 6). We also verified by immunofluorescence, using monoclonal antibodies anti-Flag (M2, Sigma F1804) and anti-HA (HA.11, Covance MMS-101P) epitopes, that each tagged remodeller was properly localized in the nucleus of ES cells.

**Verification of pluripotency in tagged ES cell line.** ES cell lines expressing a tagged remodeller were all indistinguishable in culture from their mother cell line (46C). Pluripotency of tagged ES cell lines was verified by detecting alkaline phosphatase activity on ES cell colonies 5 days after plating, using the Millipore alkaline detection kit, following manufacturer's instructions. In addition, we verified by immunofluorescence using an antibody against Oct4 (also known as Pou5f1) (Abcam ab19857, lot 943333) that expression of this pluripotency-associated transcription factor was uniform in each tagged ES cell line.

**Cell lines and ES cell culture condition.** Mouse 46C ES cells have been described previously<sup>22</sup>. 46C ES cells and their tagged derivatives were cultured at 37°C, 5% CO<sub>2</sub>, on mitomycin C-inactivated mouse embryonic fibroblasts, in DMEM (Sigma) with 15% fetal bovine serum (Invitrogen), L-glutamine (Invitrogen), MEM non-essential amino acids (Invitrogen), penicillin/streptomycin (Invitrogen),  $\beta$ -mercaptoethanol (Sigma), and a saturating amount of leukaemia inhibitory factor (LIF), as described previously<sup>23</sup>.

**Reference ES cell nucleosome map and NFR categories.** Mouse ES nucleosomal tags were acquired from a published MNase-seq data set<sup>7</sup> to make the reference map shown in Fig. 2. Reference nucleosomes were called using MACS 2.0 before assigning the first MNase-resistant nucleosome upstream and downstream of TSSs as  $-1$  and  $+1$ , respectively. Because long NFRs may actually contain MNase-sensitive nucleosome-like structures or histone-containing complexes, defining the first downstream MNase-resistant nucleosome as ' $+1$ ' is problematic, and so we refer to it as the 'first stable nucleosome'. Regions between the associated  $-1$  and  $+1$  (or first stable) nucleosomes were defined as NFRs. We further defined narrow and wide NFR categories, which have the median width of 28 bp and 808 bp, respectively. We define HFRs as lacking histones as defined by ChIP-seq.

**H3K4me3-only and bivalent gene lists.** The list of 14,623 genes used in Figs 1 and 2 was obtained by filtering all mm9 RefSeq genes<sup>24</sup>. We removed redundancies (that is, genes having the same start and end sites), unmappable genes, blacklisted genomic regions (those with artefact signal regardless of which NGS techniques were used), and genes shorter than 2 kb. The purpose of this last filtering step was to unambiguously distinguish the promoter region from the end of the genes in heat maps.

Lists of genes defined as having H3K4me3 and bivalent promoters: we first defined, among the 14,623 RefSeq genes, those with a promoter that was positive for H3K4me3 (accession number: GSM590111). This was accomplished by operating with the seqMINER platform. Tag densities from this data set were collected in a  $-500/+1,000$ -bp window around the TSS, and subjected to three successive rounds of k-means clustering, to remove all genes with a promoter that was clustered with low H3K4me3. We next conducted on this series of H3K4me3-positive promoters three successive rounds of k-means clustering, using several published data sets for H3K27me3. The genes with a promoter positive for H3K27me3 in four distinct H3K27me3 data sets (accession numbers: GSM590115, GSM590116, GSM307619 and GSM392046/GSM392047) were considered as bivalent. We eventually obtained a list of 6,481 genes with H3K4me3-only promoters, and a list of 3,411 bivalent genes.

**Tandem affinity purification of MNase-digested remodeller–nucleosome complexes.** A detailed version of this protocol is available on the protocol exchange website: <http://dx.doi.org/10.1038/protex.2014.040>. In brief, about 400 million ES cells were fixed either with formaldehyde, or with a combination of disuccinimidyl glutarate (DSG) and formaldehyde (Supplementary Table 1), then permeabilized with IGEPAL, and incubated with 2,800 units of micrococcal nuclease (MNase, New England Biolabs) in order to fragment the genome into mononucleosomes (Extended Data Fig. 1). This nucleosome preparation was next incubated with agarose beads coupled with an antibody anti-HA or anti-Flag. Anti-HA-agarose (ref. A2095) and anti-Flag-agarose (ref. A2220) beads were purchased from Sigma. After a series of washes, tagged remodeller–nucleosome complexes were eluted, either by TEV protease cleavage or by peptide competition (Supplementary Table 1). The eluted complexes were then subjected to a second immunopurification step, using beads coupled to the antibody specific of the second HA or Flag epitope. After elution, DNA was extracted from the highly purified mononucleosome fraction, and processed for high-throughput sequencing (see below). As a negative control, chromatin from untagged ES cells was subjected to the same protocol to define background signal. Two biological replicates were used for each tagged and control ES cell line, using independent cell cultures and chromatin preparations.

**High-throughput sequencing of MNase remodeller ChIP samples.** After crosslink reversion, phenol–chloroform extraction and ethanol precipitation, the DNA from remodeller–nucleosome complexes was quantified using the picogreen method (Invitrogen) or by running 1/20 of the ChIP material on a high sensitivity DNA chip on a 2100 Bioanalyzer (Agilent). Approximately 5–10 ng of ChIP DNA was used for library preparation according to the Illumina ChIP-seq protocol (ChIP-seq sample preparation kit). Following end-repair and adaptor ligation, fragments were size-selected on an agarose gel in order to purify nucleosome-sized genomic DNA fragments between 140 and 180 bp. Purified fragments were next amplified (18 cycles) and verified on a 2100 Bioanalyzer before clustering and single-read sequencing on an Illumina Genome Analyzer (GA) or GA II, according to manufacturer's instructions. Sequencing characteristics are shown in Supplementary Table 1.

**MNase remodeller ChIP-seq data analysis.** Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Ep400 and Brg1 MNase remodeller ChIP-seq short reads were mapped to mouse mm9 genome using Bowtie 0.12.7 with the followings settings:  $-a$   $-m1$   $-best$   $-strata$   $-v2$   $-p3$ . Data sets were next converted to BED format files, and data analysis was performed using the seqMINER platform<sup>25</sup> (Fig. 1c). To examine the distribution of remodellers at individual genes, we used WigMaker3 (default settings) to convert BED files into wig files, which were uploaded onto the IGV genome browser (Extended Data Fig. 2).

Nucleosome calls were made from MNase remodeller ChIP-seq tags using GeneTrack<sup>26</sup> with the following parameters:  $\sigma = 20$ , exclusion = 146. We then globally shifted tags to the median value of half distances of all nucleosome calls. GRO-seq tags<sup>10</sup> sharing the same or opposite orientation with the TSS were assigned as 'sense' and 'divergent' tags, respectively. The orientation of each NFR was arranged so that sense transcription proceeds to the right. ES nucleosomal tags, globally shifted tags from MNase remodeller ChIP-seq (this current study), tags from DHS regions (Mouse ENCODE), GRO-seq oriented tags from transcriptionally engaged Pol II and CpG islands (UCSC, mm9 build) were then aligned to the midpoint of each NFR. Promoter regions were then sorted by NFR length and visualized by Java TreeView (Fig. 2a, b).

CpG island information was retrieved from UCSC (mm9 build) and assigned to the closest TSS by using bedtools. We noticed that promoters with wide NFRs were mostly CpG island (CpGI)-rich, while those with narrow NFRs were globally CpGI-poor, in agreement with a previous report showing that CpGIs induce nucleosome exclusion<sup>9</sup> (Fig. 2b).

Tags from reference nucleosomes<sup>7</sup>, remodeller-interacting nucleosomes (this study) and transcriptionally engaged Pol II (GRO-seq)<sup>10</sup> were aligned to nucleosome  $-1$  and  $+1$  (or the first stable nucleosome) dyad positions. The direction of each dyad was assigned according to the orientation of its associated TSS, the orientation of which was arranged so that the transcription proceeds to the right. After normalization to the gene count in the two different NFR subclasses, tags were plotted from the NFR midpoint to 500 bp distal to the reference nucleosome. An  $x$  axis gap in the NFR was introduced to normalize variations in NFR length inside each class.

**Pearson correlation analysis.** We used DNaseI-Seq data from the mouse ENCODE consortium (GSM1004653) for the identification of DHS regions in the mouse ES cell genome. DHS regions were defined using MACS 2.0 (ref. 27) (default setting), which resulted in the identification of 139,454 DHS regions. Each of these DHS regions was represented as a 500-bp window ( $-250$  bp/ $+250$  bp) centred on the midpoint of the DHS peak. DHS regions overlapping with the blacklisted (high background signal) genomic areas (mm9) were removed, resulting in a final list of 138,582 DHS regions. Tags from each tested ChIP-seq data set were summed up for

each DHS region before pair-wise Pearson correlation comparison. The  $R^2$  value from each pair-wise Pearson correlation was then visualized by heat map (Fig. 1a).

Pearson correlation analysis at promoter-like DHS regions. Operating with the seqMINER platform, we retrieved, from the 138,582 DHS regions list, those positive for H3K4me3, TBP and Pol II S5ph. We obtained 16,300 promoter-like DHS regions befitting the criteria. Pair-wise Pearson correlation was performed and plotted (Fig. 1b) as described for Fig. 1a.

**RNA preparation from ES cells depleted of each remodeller by shRNA.** We used the pHYPER shRNA vector for remodeller depletion in ES cells, as previously described<sup>28</sup>. shRNA design was performed using DSIR software (<http://biodev.extra.cea.fr/DSIR/DSIR.html>). Below are the shRNAs selected for each remodeller. The sense strand sequence is given; the rest of the shRNA sequence is as described previously<sup>28</sup>. *Chd1* shRNA 1: 5'-GCAAAGACGGCGACTAGAAGA-3'; *Chd1* shRNA 2: 5'-GACAGTGCTTAATCAAGATCG-3'; *Chd4* shRNA 1: 5'-GGACGACGATTAGATGTAGA-3'; *Chd4* shRNA 2: 5'-GCTGACGTCTTC AAGAATATG-3'; *Chd6* shRNA 1: 5'-GTACTATCTGCTATCCTAGA-3'; *Chd6* shRNA 2: 5'-CAGTCAGAACCCACAATAACT-3'; *Chd8* shRNA 1: 5'-GCAGT TACACTGACGTCTACA-3'; *Chd8* shRNA 2: 5'-GACTTCTGTGTA CCGCTCAAGA-3'; *Chd9* shRNA 1: 5'-TATACCAATTGAACAAGAGCC-3'; *Chd9* shRNA 2: 5'-AGTTAAAGTCTACAGATTAGT-3'; *Ep400* shRNA 1: 5'-GGTAAAGAGTCCAGATTAAAG-3'; *Ep400* shRNA 2: 5'-GGTCCAC ACTCAACAACGAGC-3'; *Smarca4* shRNA 1: 5'-ACTTCTTGATAGAA TTCTACC-3'; *Smarca4* shRNA 2: 5'-CCTTCGAACAGTGGTTCAATG-3'.

Each shRNA was transfected in its corresponding tagged ES cell line, to follow remodeller depletion by western blotting using monoclonal antibodies anti-Flag (M2, Sigma F1804), or anti-HA (H7, Sigma H3663) epitopes (Extended Data Fig. 6), in comparison with the signal obtained with a control antibody anti-Gapdh (Abcam ab9485).

The pHYPER shRNA vectors were transfected in ES cell by electroporation, using an Amaxa nucleofector (Lonza). Twenty-four hours after transfection, puromycin ( $2 \mu\text{g ml}^{-1}$ ) selection was applied for an additional 48 h period, before cell collection and RNA preparation, except for *Chd4*, for which cells were collected after 30 h of selection. Total RNA was extracted using an RNeasy kit (Qiagen). Total RNA yield was determined using a NanoDrop ND-100 (Labtech). Total RNA profiles were recorded using a Bioanalyzer 2100 (Agilent). For each remodeller, RNA was prepared from three independent transfection experiments, and processed for transcriptome analysis.

**Analysis of gene expression in 46C ES cells by RNA-seq.** 46C ES cells were amplified on feeder cells except for the last passage, at which point cells were plated onto 60-mm dishes coated with gelatine, and grown to 70% confluence in D15 medium with LIF. Total RNA was extracted using an RNeasy Kit (Qiagen). The RNA quality was verified on a 2100 Bioanalyzer. Library preparation was performed using the Illumina mRNAseq sample preparation kit according to manufacturer's instructions. Briefly, the total RNA was depleted of ribosomal RNA using the Sera-mag Magnetic Oligo (dT) Beads (Illumina) and after mRNA fragmentation, reverse transcription and second strand cDNA synthesis the Illumina specific adaptors were ligated. The ligation product was then purified and enriched with 15 cycles of PCR to create the final library for single-read sequencing of 75 bp carried out on an Illumina GAIIX.

To keep only sequences of good quality, we retained the first 40 bp of each read and discarded all sequences with more than 10% of bases having a quality score below 20, using FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Mapping of these sequences onto the mm9 assembly of mouse genome and RPKM computation were then performed using ERANGE v3.1.0 (ref. 29) and bowtie v0.12.0 (ref. 30). In brief, a splice file was created with UCSC known genes and maxBorder = 36. We created an expanded genome containing genomic and splice-spanning sequences using bowtie-build and bowtie was used to map the reads onto this expanded genome. Then the ERANGE runStandardAnalysis.sh script was used to compute RPKM values following steps previously described<sup>29</sup>, using a consolidation radius of 20 kb.

**qRT-PCR analysis of gene expression.** Random-primed reverse transcription was performed at 52 °C in 20  $\mu\text{l}$  using Maxima First strand cDNA synthesis kit (Thermo Scientific) with 1  $\mu\text{g}$  of total RNA isolated from ES cells (Qiagen), quantified with NanoDrop instrument (Thermo Scientific). Reverse transcription products were diluted 40-fold before use. Composition of quantitative PCR assay included 2.5  $\mu\text{l}$  of the diluted RT reaction, 0.2–0.5 mM forward and reverse primers, and 1  $\times$  Maxima SYBR Green qPCR Master Mix (Thermo Scientific). Reactions were performed in a 10  $\mu\text{l}$  total volume. Amplification was performed as follows: 2 min at 95 °C, 40 cycles at 95 °C for 15 s and 60 °C for 60 s in the ABI/Prism 7900HT real-time PCR machine (Applied Biosystems). The real-time fluorescent data from qPCR were analysed with the Sequence Detection System 2.3 (Applied Biosystems). Each qPCR reaction was performed using the set of primer pairs listed in Supplementary Table 2, validated for their specificity and efficiency of amplification. All reactions

were performed in triplicates, using RNA prepared from three independent cell transfection experiments. Control reactions without enzyme were verified to be negative. Relative expression was calculated after normalization with three reference genes (*Actb*, *Nmt1* and *Ddb1*), validated for this study.

**Transcriptome analysis in remodeller-depleted ES cells.** cRNA was synthesized, amplified and purified using the Illumina TotalPrep RNA Amplification Kit (Life Technologies) following Manufacturer's instructions. In brief, 200 ng of RNA were used to prepare double-stranded cDNA using a T7 oligonucleotide (dT) primer. Second-strand synthesis was followed by *in vitro* transcription in the presence of biotinylated nucleotides. cRNA samples were hybridized to the Illumina BeadChips Mouse WG-6v2.0 arrays. These BeadChips contain 45,281 unique 50-mer oligonucleotides in total, with hybridization to each probe assessed at 30 different beads on average. A total of 26,822 probes (59%) are targeted at RefSeq transcripts, and the remaining 18,459 (41%) are for other transcripts. BeadChips were scanned on the Illumina iScan scanner using Illumina BeadScan image data acquisition software (version 2.3). Data were then normalized using the 'normalize quantiles' function in the GenomeStudio Software (version 1.9.0). Following analyses were done using Genespring software (version 13.0-GX).

For Brg1, we used a previously published transcriptome data set, in which loss of Brg1 function was obtained by genetic ablation<sup>18</sup>. All array analyses were undertaken using the Limma package from the R/Bioconductor software (R-Development-Core-Team, 2007). Microarray spot intensities were normalized using the RMA method as implemented in the R affy package. Normalized measures served to compute the  $\log_2$  ratios for each gene between the wild-type strain and the Brg1 knockout mutant. Then, to identify genes with a  $\log_2$  ratio significantly different between the mutant and wild-type strain,  $P$  values were calculated for each gene using a moderated  $t$ -test. The moderated  $t$ -test applied here was based on an empirical Bayes analysis and was equivalent to shrinkage (or expansion) of the estimated sample variances towards a pooled estimate, resulting in a more stable inference. Finally, adjusted  $P$  values were calculated using the false discovery rate (FDR)-controlling procedure of Benjamini and Hochberg.

**Analysis of gene deregulation.** We identified deregulated genes using the thresholds of 0.05 for the  $P$  value, and 1.5 for the fold change (FC 1.5). This FC 1.5 threshold was chosen based on a previous study on Brg1 (ref. 18), and also because it was compatible with the analysis of the remodellers more modestly involved in transcriptional control in ES cells such as *Chd1*, *Chd6* and *Chd8*. Note that seemingly modest fold changes might arise from many sources including a response lag, residual remodelling activity, and relatively high experimental background. Using a FC 2 threshold, we could, however, confirm that *Ep400*, *Chd4* and *Brg1* are important transcriptional regulators in ES cells, with 535, 293 and 570 genes deregulated, respectively. This level of deregulation is indicative of a context-specific function of remodellers in transcriptional activation or repression, which is distinct from the function of general transcription factors, whose depletion is expected to affect most genes.

Statistical analysis of the differences in transcriptional activation and repression by remodellers was performed using a two-sample test for equality of proportions with continuity correction.

For the generation of GC-content-based lists of promoters, we used the list of promoters defined in figure 3 of ref. 15, which we crossed with the 14,623 promoter list, to obtain a list of 6,317 promoters rank ordered according to GC content.

In Fig. 3b, we compared the percentages of genes either down- or upregulated by loss of function of each remodeller in the following two groups: (1) NFR length classes: genes from the narrow and wide NFR classes shown in Fig. 2a were each further divided into two subclasses, which resulted in the following four categories: narrow NFR subclass 1 (NFR < 15 bp), narrow NFR subclass 2 (15–115 bp NFR), wide NFR subclass 1 (116–504 bp) and wide NFR subclass 2 (505–1,500 bp). Genes in these groups were further subdivided into H3K4me3 and bivalent subgroups. (2) GC content classes: genes were divided into four quartiles based on GC content at promoters and further subdivided into H3K4me3 and bivalent subclasses. The number of genes analysed in Fig. 3b is indicated in brackets for the following subgroups. H3K4me3 genes: narrow NFR subclass 1 (739), subclass 2 (1,829), wide NFR subclass 1 (2,613), subclass 2 (1,253), GC content quartile 1 (low GC content) (450), quartile 2 (719), quartile 3 (644), quartile 4 (high GC content) (430). Bivalent genes: narrow NFR subclass 1 (271), subclass 2 (866), wide NFR subclass 1 (2,266), subclass 2 (1,184), GC content quartile 1 (220), quartile 2 (485), quartile 3 (750) and quartile 4 (1149).

**FAIRE-seq.** FAIRE was performed as described<sup>31</sup> with modifications. 46C ES cells were amplified as described above for RNA preparation. Formaldehyde was added directly to the growth media (final concentration 1%), and cells were fixed for 5 min at room temperature. After quenching with glycine (125 mM) and several washes, cells were collected, resuspended in 500  $\mu\text{l}$  of cold lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl, pH 8.0 and 1 mM EDTA) and disrupted using glass beads for five 1-min sessions with 2-min incubations on ice



between disruption sessions. Samples were then sonicated for 16 sessions of 1 min (30 s on/30 s off) using a bioruptor (Diagenode) at max intensity, at 4 °C. After centrifugation, the supernatant was extracted twice with phenol–chloroform. The aqueous fractions were collected and pooled, and a final phenol–chloroform extraction was performed before DNA precipitation. FAIRE experiments were realized in triplicate, using independent ES cell cultures. Before sequencing, FAIRE DNA was analysed and quantified by running 1/25 of the FAIRE material on a high sensitivity DNA chip on a 2100 Bioanalyzer (Agilent, USA). Approximately 20 ng of FAIRE DNA was used for library preparation according to manufacturer's instructions using the ChIP-seq sample preparation kit (Illumina). Single-read sequencing (36 bp) was performed on a Genome Analyzer II (Illumina).

**ATAC-seq.** ES cells were grown and transfected with shRNA vectors as described for RNA analysis. Biological replicates were obtained by performing two independent transfection experiments for each shRNA vector. ATAC-seq libraries were constructed by adapting a published protocol<sup>20</sup>. In brief, 50,000 cells were collected, washed with cold PBS and resuspended in 50 µl of ES buffer (10 mM Tris, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>). Permeabilized cells were resuspended in 50 µl transposase reaction (1 × tagmentation buffer, 1.0–1.5 µl Tn5 transposase enzyme (Illumina)) and incubated for 30 min at 37 °C. Subsequent steps of the protocol were performed as previously described<sup>20</sup>. Libraries were purified using a Qiagen MinElute kit and Ampure XP magnetic beads (1:1.6 ratio) to remove remaining adapters. Libraries were controlled using a 2100 Bioanalyzer, and an aliquot of each library was sequenced at low depth onto a MiSeq platform to control duplicate level and estimate DNA concentration. Each library was then paired-end sequenced (2 × 100 bp) on a HiSeq instrument (Illumina).

**Analysis of ATAC-seq data.** As ATAC-seq libraries are composed in large part of short genomic DNA fragments, reads were cropped to 50 bp using trimmomatic-0.32 to optimize paired-end alignment. Reads were aligned to the mouse genome (mm9) using Bowtie with the parameters -m1-best-strata -X2000, with two mismatches permitted in the seed (default value). The -X2000 option allows the fragments <2 kb to align and -m1 parameter keeps only unique aligning reads. Duplicated reads were removed with picard-tools-1.85. To perform differential analysis, libraries were adjusted to 33 million aligned reads using samtools-1.2 and by making a random permutation of initial input libraries (shuf linux command line). Adjusted BAM data sets were next converted to BED. We used the seqMINER platform with the lists of 6,481 H3K4me3-only and 3,411 bivalent genes described above, to collect tag densities from ATAC-seq data sets, in a window of -2 kb/+2 kb around the TSS. Output tag density files were analysed using R software to establish average ATAC-seq signal profiles shown in Extended Data Fig. 8.

**MNase-seq following remodeller depletion in ES cells.** ES cells were grown and transfected with shRNA vectors as described above. Biological replicates were obtained by performing two independent transfection experiments for each shRNA vector. For each experiment, 1 million cells were fixed 10 min in ES cell culture medium containing 1% formaldehyde, quenched with glycine (125 mM), washed with PBS buffer, collected in 175 µl of solution I (15 mM Tris-HCl, pH 7.5, 0.3 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl<sub>2</sub> and 0.1 mM EGTA), and stored on ice. Cells were permeabilized by adding 175 µl of solution II (solution I with 0.8% Igepal CA-630 (Sigma)) and incubating for 15 min on ice. We next added 700 µl of MNase digestion buffer (50 mM Tris-HCl, pH 7.5, 0.3 M sucrose, 15 mM KCl, 60 mM NaCl, 4 mM MgCl<sub>2</sub> and 2 mM CaCl<sub>2</sub>), 4 U of MNase, and incubated for 10 min at 37 °C. MNase digestion was stopped by adding 10 mM EDTA (final concentration), and storing on ice. Cells were then disrupted by 15 passages through a 25 G needle, followed by a 10 min centrifugation at 18,000g. The supernatant was collected and incubated for 1 h at 65 °C with 15 µg of RNase A. We next added 10 µg of proteinase K, adjusted each sample to 0.1% SDS (final concentration) and incubated for 2 h at 55 °C. NaCl concentration was then adjusted to 200 mM and the samples were incubated overnight at 65 °C for crosslink reversal. DNA was purified from each sample by phenol–chloroform extraction followed by ethanol precipitation.

Purified DNA (20 ng) was used for library preparation according to manufacturer's instructions, using Ultralow ovation library system (Nugen). Following end-repair and adaptor ligation, fragments were size-selected onto an agarose gel in order to purify genomic DNA fragments between ~60 and 220 bp. Libraries were verified using a 2100 Bioanalyzer before clustering and paired-read sequencing. Sequencing of each sample was performed in a single lane of a HiSeq instrument (Illumina).

**Analysis of MNase-seq data.** The midpoint of each paired-end sequencing read was used to represent dyad location of each nucleosomal tag. We assumed that remodeller depletion has no bulk effect on nucleosome occupancy, hence the total reads of control and remodeller-depleted cells were adjusted to be the same. The adjusted tags were aligned to -1 nucleosome dyads

(determined by the first MNase-defined peak upstream of annotated RefSeq TSS), or the first stable (MNase-defined) nucleosome dyad position downstream of the TSS for different NFR categories. These tags were further normalized to the amount of genes involved in each NFR class. The normalized tags were then binned (5 bp) and smoothed (10-bin moving average) before plotting (Fig. 3c). Distances (bp) are indicated relative to these reference points. An x axis gap in the NFR was introduced to normalize variations in NFR length inside each class.

**Pol II ChIP-exo.** ES cells were grown and transfected with shRNA vectors as described above. Biological replicates were obtained by performing two independent transfection experiments for each shRNA vector. Following a 10 min fixation with 1% formaldehyde in ES cell culture medium, chromatin was prepared from 5–10 million cells and sonicated as described<sup>32</sup>. ChIP-exo experiments were carried out essentially as described<sup>33</sup>. This included an immunoprecipitation step using antibodies against Pol II (sc-899, Santa Cruz Biotechnology) attached to magnetic beads, followed by DNA polishing, A-tailing, Illumina adaptor ligation (ExA2), and lambda and recJ exonuclease digestion on the beads. After elution, a primer was annealed to EXA2 and extended with phi29 DNA polymerase, then A-tailed. A second Illumina adaptor was then ligated, and the products PCR-amplified and gel-purified. Sequencing was performed using NextSeq500. Uniquely aligned sequence tags were mapped to the mouse genome (mm9) using BWA-MEM (version 0.7.9a-r786)<sup>34</sup>. The uniquely aligned sequence tags were used for the downstream analysis.

**Analysis of Pol II ChIP-exo data.** The 5' end of mapped tags, representing exonuclease stop sites, were consolidated into peak calls (sigma = 5, exclusion = 20) using GeneTrack<sup>26</sup>, and peak pairs were matched when found on opposite strands and 0–100 bp apart in the 3' direction. Tags were globally shifted to the median value of half distance between all peak pairs.

These global shifted tags were then aligned relative to the annotated RefSeq TSSs for H3K4me3-only and bivalent promoters separately before further carved out remodeller-affected genes. We assumed that having remodeller deletion bore no bulk change on Pol II occupancy, and hence total tags among wild type and all remodeller mutants were normalized to be the same. To make direct comparison between different gene groups, we further normalized tags to the amount of genes within the group. These normalized tags were then smoothed (5 bp binned before 10-bin moving average) before plotting (Extended Data Fig. 9a).

To examine Pol II occupancy change in remodeller mutants among different promoter groups, we first calculated total Pol II occupancy by summing up tags from transcript start to end sites (annotated RefSeq TSS and TES, respectively<sup>24</sup>) for the tested genes. Change in Pol II occupancy was calculated by dividing the total Pol II occupancy of mutant by that of wild type before log<sub>2</sub> transformation and bargraph plotting (Extended Data Fig. 9b).

**Average binding profiles (Extended Data Fig. 3).** Genes were rank-ordered according to reads per kb of transcript per million mapped reads (rpkm) and divided in four quartiles (highest: Q4, second: Q3, third: Q2 and lowest: Q1). Operating with the k-means clustering function of seqMINER, genes in each quartile were further subdivided in H3K4me3-only and bivalent genes, as described above.

Using these lists of genes, tag densities from remodeller ChIP-seq data sets were collected in a window of -2 kb/+2 kb around the TSS, except for Chd2, for which densities were collected from the TSS until +4 kb. Output tag density files were first analysed using R software to establish average binding profiles. Statistical comparisons were performed between remodeller distributions at H3K4me3 promoters, to assess a significant increasing trend among distributions. Differences between successive pairs of quartiles (Q4 – Q3, Q3 – Q2 and Q2 – Q1) were compared against a null distribution using a one side *t*-test.

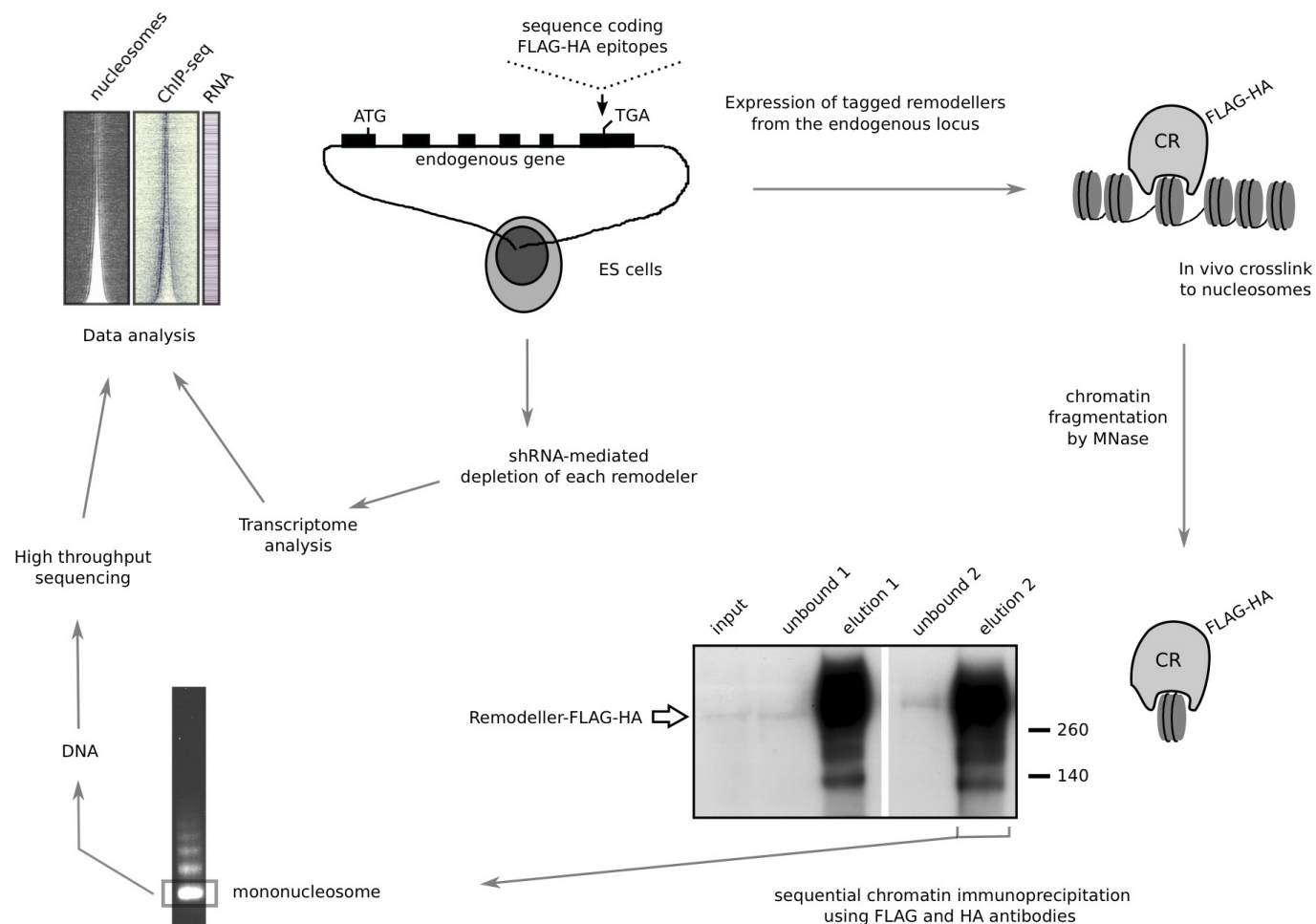
The respective *P* values are reported for each remodeller: Chd1, Q4 – Q3  $P = 1.371138 \times 10^{-27}$ ; Q3 – Q2  $P = 1.728126 \times 10^{-16}$ ; Q2 – Q1  $P = 7.985217 \times 10^{-23}$ ; Chd2, Q4 – Q3  $P = 7.543473 \times 10^{-33}$ ; Q3 – Q2  $P = 1.115223 \times 10^{-25}$ ; Q2 – Q1  $P = 3.283427 \times 10^{-38}$ ; Chd4, Q4 – Q3  $P = 0.2094255$ ; Q3 – Q2  $P = 0.1081455$ ; Q2 – Q1  $P = 0.07202865$ ; Chd6, Q4 – Q3  $P = 0.4168748$ ; Q3 – Q2  $P = 0.1534144$ ; Q2 – Q1  $P = 0.01138035$ ; Chd8, Q4 – Q3  $P = 4.031959 \times 10^{-15}$ ; Q3 – Q2  $P = 1.231527 \times 10^{-6}$ ; Q2 – Q1  $P = 1.34455 \times 10^{-9}$ ; Chd9, Q4 – Q3  $P = 9.484578 \times 10^{-44}$ ; Q3 – Q2  $P = 1.059783 \times 10^{-14}$ ; Q2 – Q1  $P = 4.646352 \times 10^{-28}$ ; Ep400, Q4 – Q3  $P = 3.046796 \times 10^{-20}$ ; Q3 – Q2  $P = 1.215304 \times 10^{-14}$ ; Q2 – Q1  $P = 6.462667 \times 10^{-11}$ ; Brg1, Q4 – Q3  $P = 3.512021 \times 10^{-24}$ ; Q3 – Q2  $P = 2.515217 \times 10^{-7}$ ; Q2 – Q1  $P = 0.977422$ .

We concluded from this analysis that Chd1, Chd2, Chd9 and Ep400 binding at promoters is tightly linked to gene expression level. By contrast, Brg1, Chd4 and Chd6 deposition showed little correlation with gene expression level (statistical test failed for at least one comparison for these remodelers). While statistical analysis of Chd8 distributions concluded to significant differences between quartiles, inspection of distributions in Extended Data Fig. 3 showed that Chd8 binding profile was intermediate between these two categories.



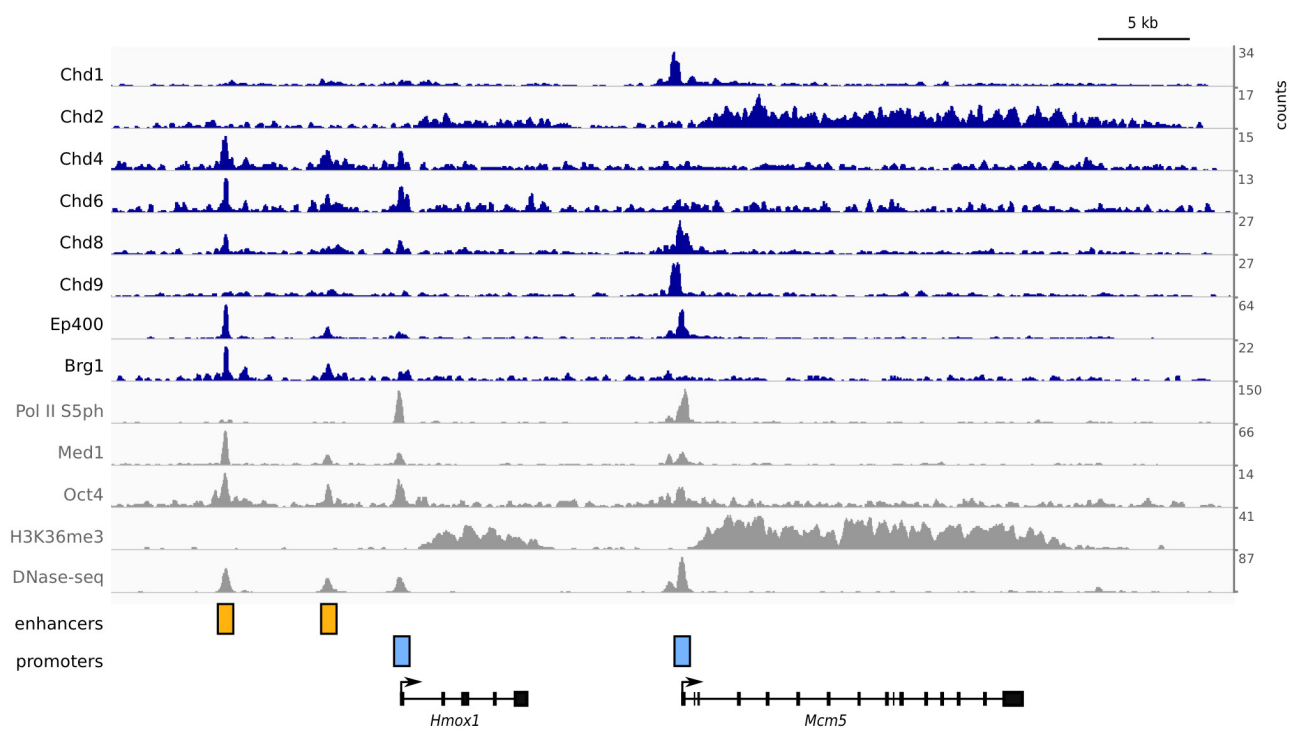
**Accession numbers and references of the publicly available data sets used in Figs 1 and 2 and Extended Data Figs 1, 4 and 7.** Brg1 (ref. 35): GSM359413; DNase-seq: GSM1014154; Ezh2 (ref. 15): GSM590132; GRO-seq<sup>10</sup>: GSM665994; H2A.Z<sup>13,14</sup>: GSM958501, DRP001103; H3.3 (ref. 12): GSM1386359; H3K27ac (ref. 16): GSM594578; H3K27me3 (ref. 15): GSM590115; H3K36me3 (ref. 15): GSM590119; H3K4me3 (ref. 15): GSM590111; Med1 (ref. 36): GSM560347; Mi2b (Chd4)<sup>8</sup>: GSM687284; MNase-seq<sup>7</sup>: GSM1004653; Oct4, Sox2, Nanog: GSM1082340; Pol II S5ph (ref. 37): GSM515662; TBP: GSM958503.

21. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476–484 (2003).
22. Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature Biotechnol.* **21**, 183–186 (2003).
23. Tessarollo, L. Manipulating mouse embryonic stem cells. *Methods Mol. Biol.* **158**, 47–63 (2001).
24. Pruitt, K. D. et al. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* **42**, D756–D763 (2014).
25. Ye, T. et al. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* **39**, e35 (2011).
26. Albert, I., Wachi, S., Jiang, C. & Pugh, B. F. GeneTrack—a genomic data processing and visualization framework. *Bioinformatics* **24**, 1305–1306 (2008).
27. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nature Protocols* **7**, 1728–1740 (2012).
28. Berlivet, S., Houliard, M. & Gérard, M. Loss-of-function studies in mouse embryonic stem cells using the pHYPER shRNA plasmid vector. *Methods Mol. Biol.* **650**, 85–100 (2010).
29. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
30. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
31. Giresi, P. G. & Lieb, J. D. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods* **48**, 233–239 (2009).
32. Carrière, L. et al. Genomic binding of Pol III transcription machinery and relationship with TFIS transcription factor distribution in mouse embryonic stem cells. *Nucleic Acids Res.* **40**, 270–283 (2012).
33. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
34. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
35. Ho, L. et al. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc. Natl Acad. Sci. USA* **106**, 5187–5191 (2009).
36. Kagey, M. H. et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
37. Rahl, P. B. et al. c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).

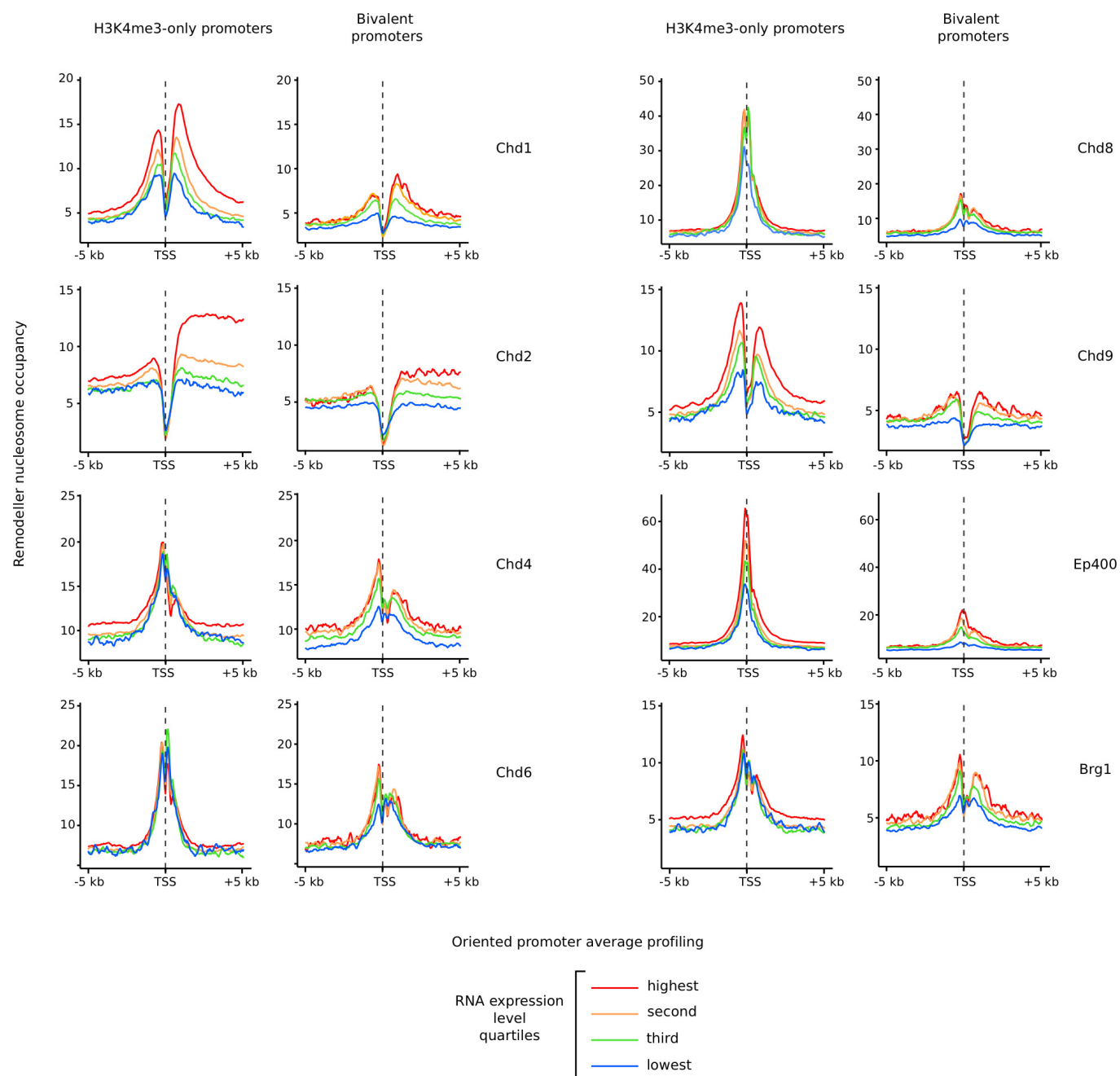


**Extended Data Figure 1 | Experimental strategy for genome-wide remodeller-nucleosome interactions and transcriptome analysis in ES cells.** Using homologous recombination in ES cells, a sequence encoding a combination of Flag and haemagglutinin (HA) epitopes was introduced at the 3' end of the coding sequence of the genes encoding the catalytic subunit of each remodeller. After *in vivo* crosslinking, chromatin was prepared and fragmented to mononucleosomes by MNase. Remodeller-bound

mononucleosomes were isolated using a double-immunoaffinity procedure. Immunopurification efficiency was assessed by western blotting. Deep sequencing of the DNA from purified nucleosomes allowed the mapping of remodeller-bound nucleosomes across the mouse genome. The same tagged ES cell lines were used for shRNA-mediated depletion of remodellers and transcriptome analysis.



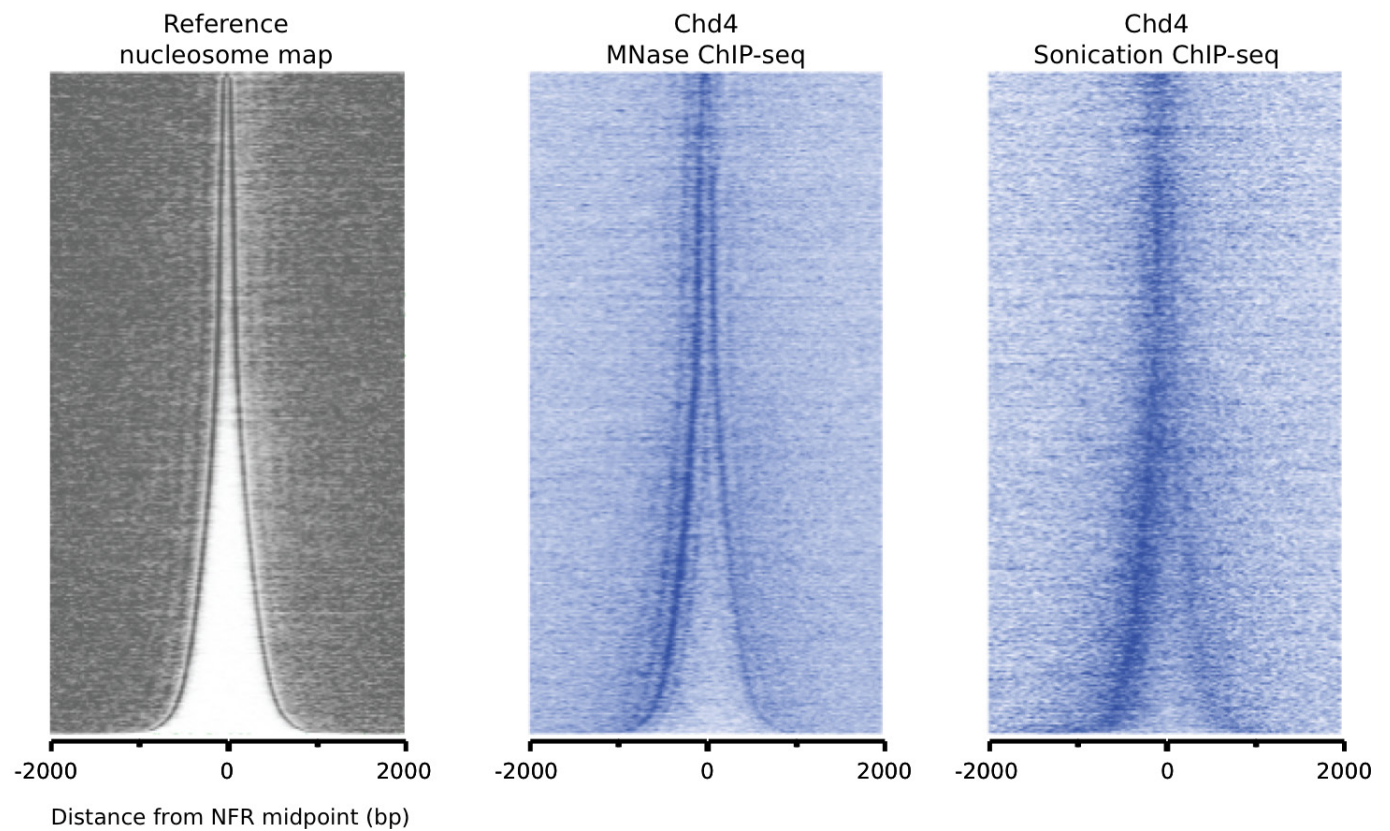
**Extended Data Figure 2 | Remodeller binding profile at a representative locus.** Counts indicate reads per 10 million. Promoters and enhancers are highlighted by blue and orange squares, respectively.



**Extended Data Figure 3 | Relationship between remodeller enrichment at promoters and RNA expression level.** Average binding profile of remodellers at promoters, divided in four quartiles based on RNA expression level of the corresponding genes. All promoters are transcribed from left to right. Promoter binding intensity of Chd1, Chd2, Chd9 and Ep400 at H3K4me3 promoters was correlated with RNA expression

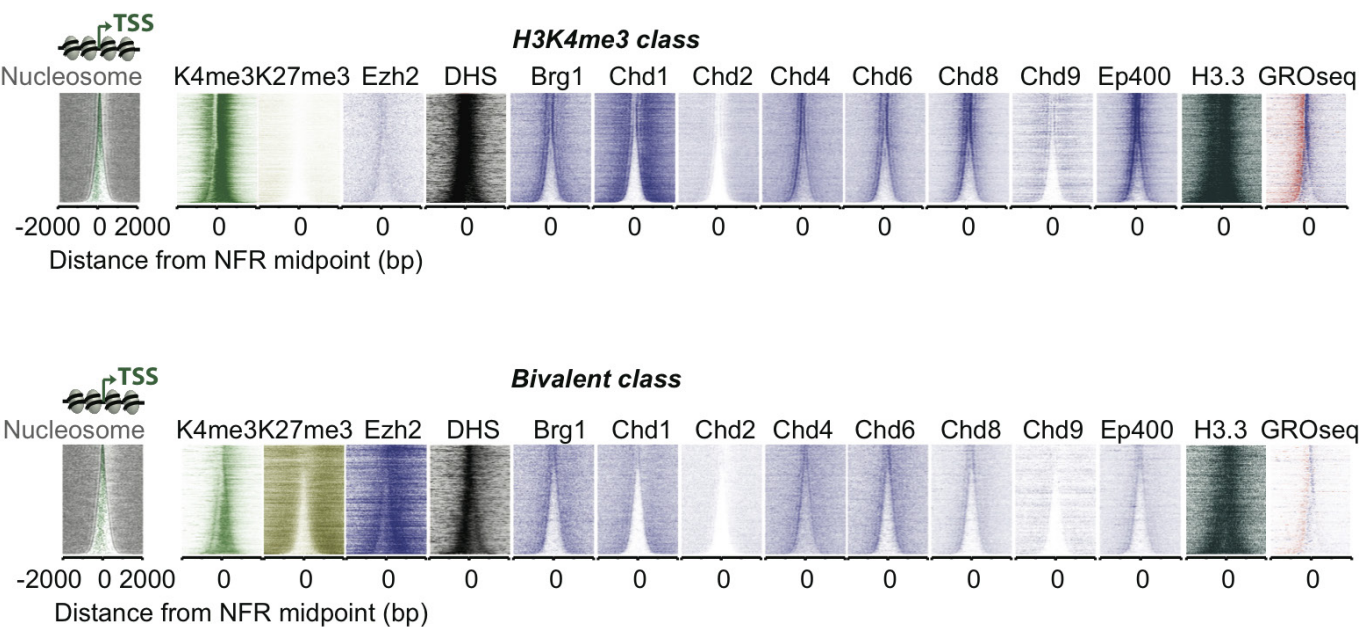
(see Methods). Consequently, binding of these remodellers to bivalent promoters, which are transcribed at lower levels, showed a significant reduction compared to H3K4me3 promoters. By contrast, Chd4, Chd6 and Brg1 enrichment at promoters showed little correlation with the transcription level of the corresponding genes, and was only slightly lower at bivalent, compared to H3K4me3 promoters.



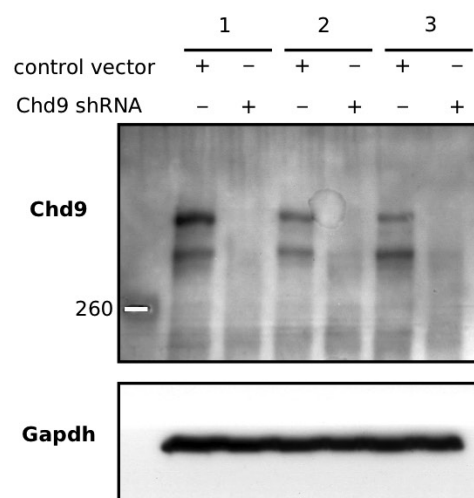
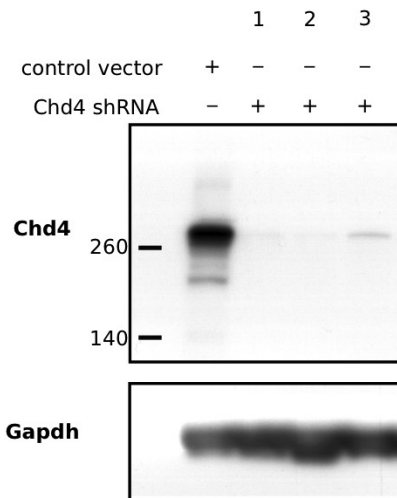
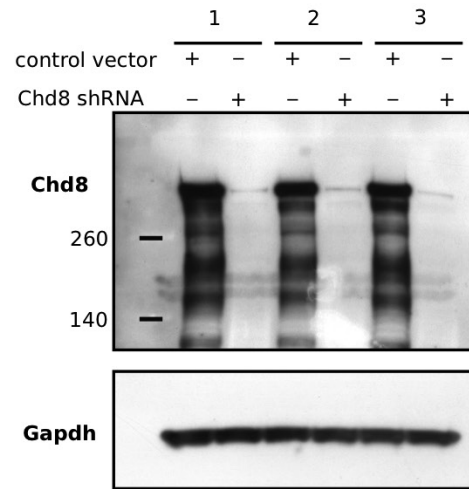
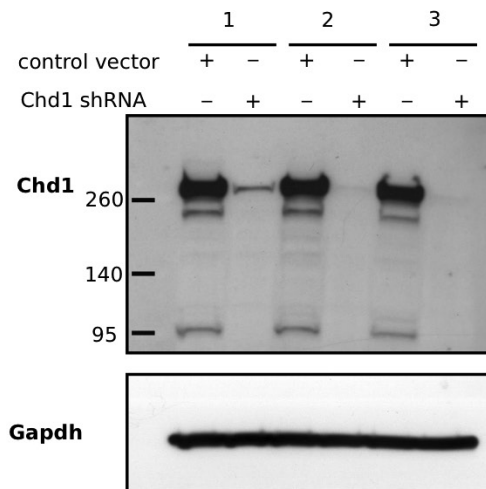
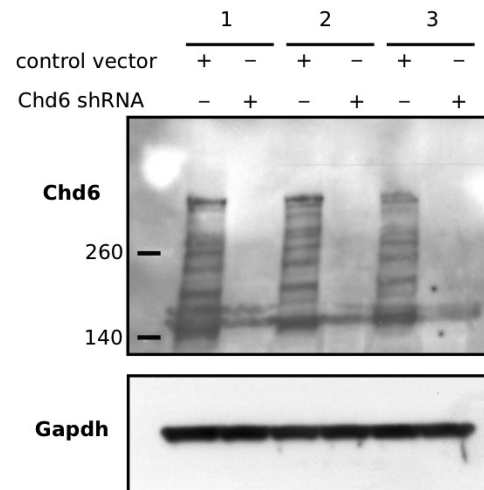
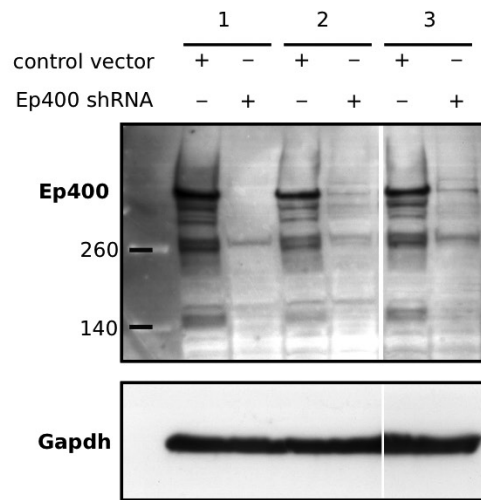


**Extended Data Figure 4 | Comparison of MNase ChIP-seq and sonication ChIP-seq for Chd4.** The left panel shows the reference nucleosome map of 14,623 RefSeq genes, rank-ordered from smallest to largest NFR length, as in Fig. 2. The two panels on the right compare

the distribution patterns obtained for Chd4 either by MNase ChIP-seq, with chromatin prepared from Chd4-tagged ES cells, or by ChIP-seq with sonicated chromatin (data set accession number: GSM687284).

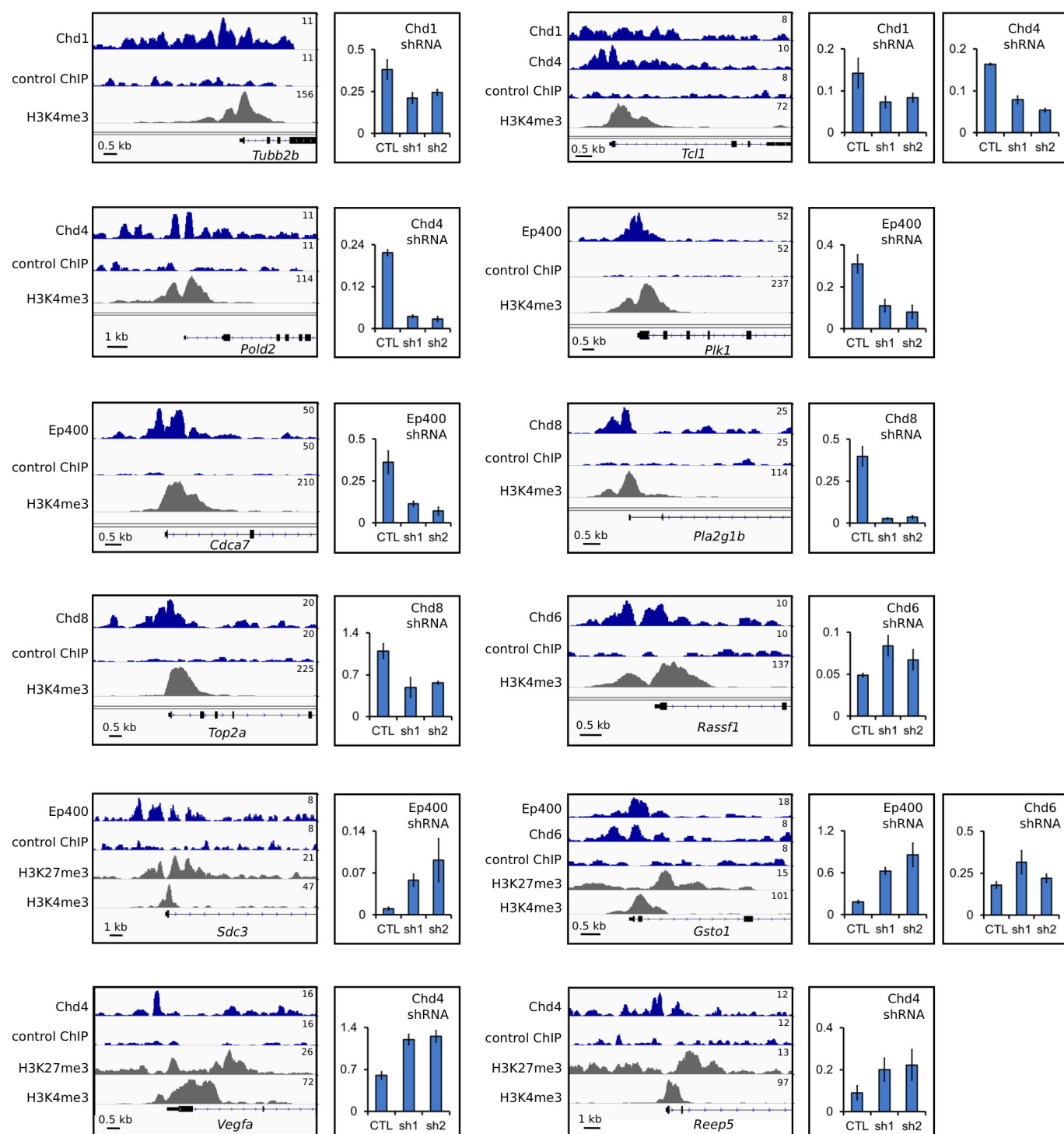


**Extended Data Figure 5 | Nucleosome targeting by remodellers at H3K4me3-only and bivalent promoters.** Remodeller-bound nucleosomal tags were aligned to the promoters of 6,481 active (H3K4me3 promoters) or 3,411 bivalent genes, rank-ordered from narrow to wide NFR. Corresponding reference nucleosomes, remodeller occupancy and the other indicated features are shown as in Fig. 2.



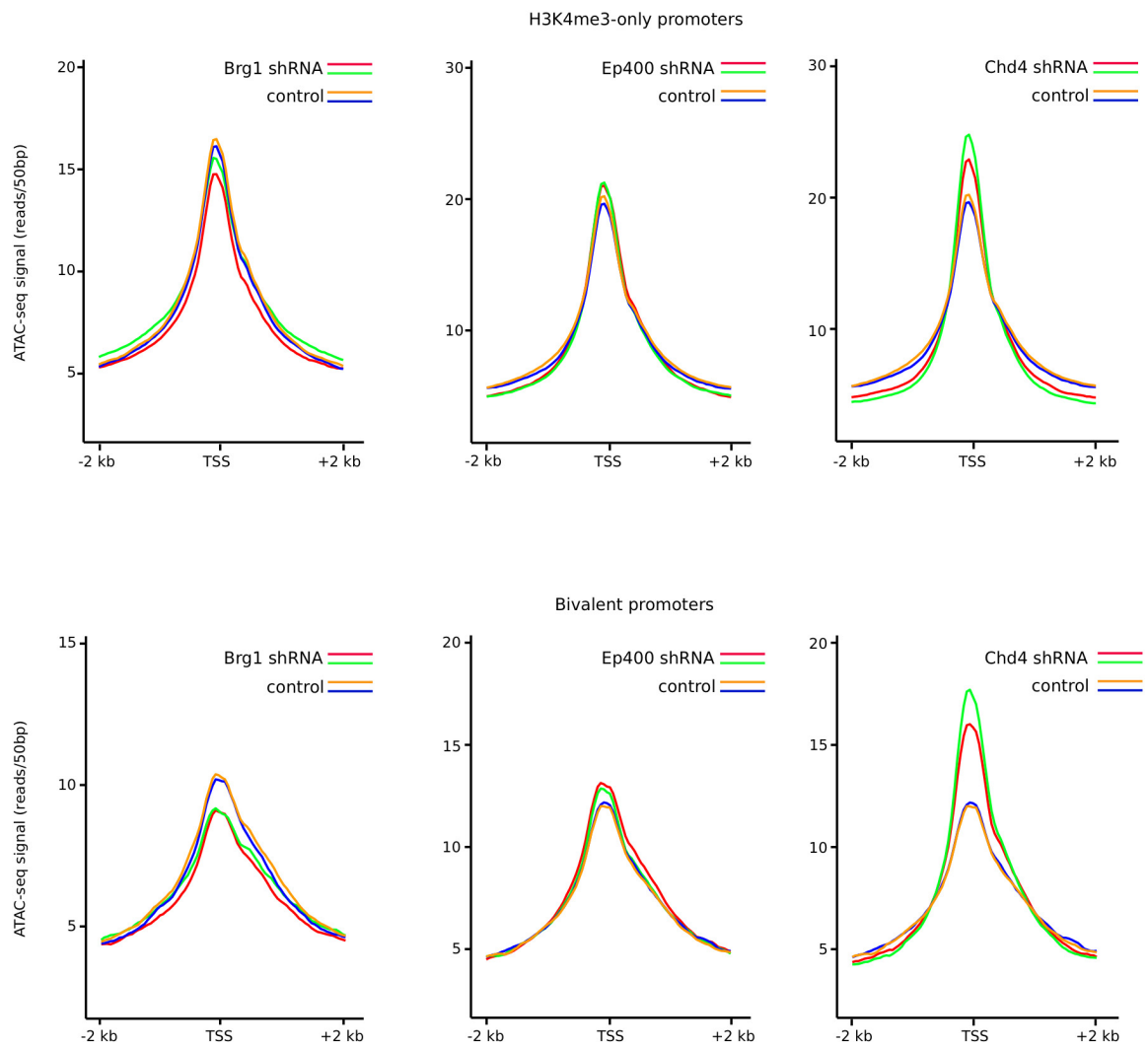
**Extended Data Figure 6 | Western blot analysis of remodeller depletion by shRNA for transcriptome analysis.** ES cells tagged for each remodeller were transfected with the corresponding shRNA vector, or a control plasmid. After puromycin selection, ES cells were collected for RNA

preparation and western blot analysis. Three independent experiments (indicated as 1, 2 and 3) were performed for each remodeller. Remodeller depletion was assessed using antibodies against Flag or HA epitopes. Loading control: Gapdh. For gel source data, see Supplementary Fig. 1.



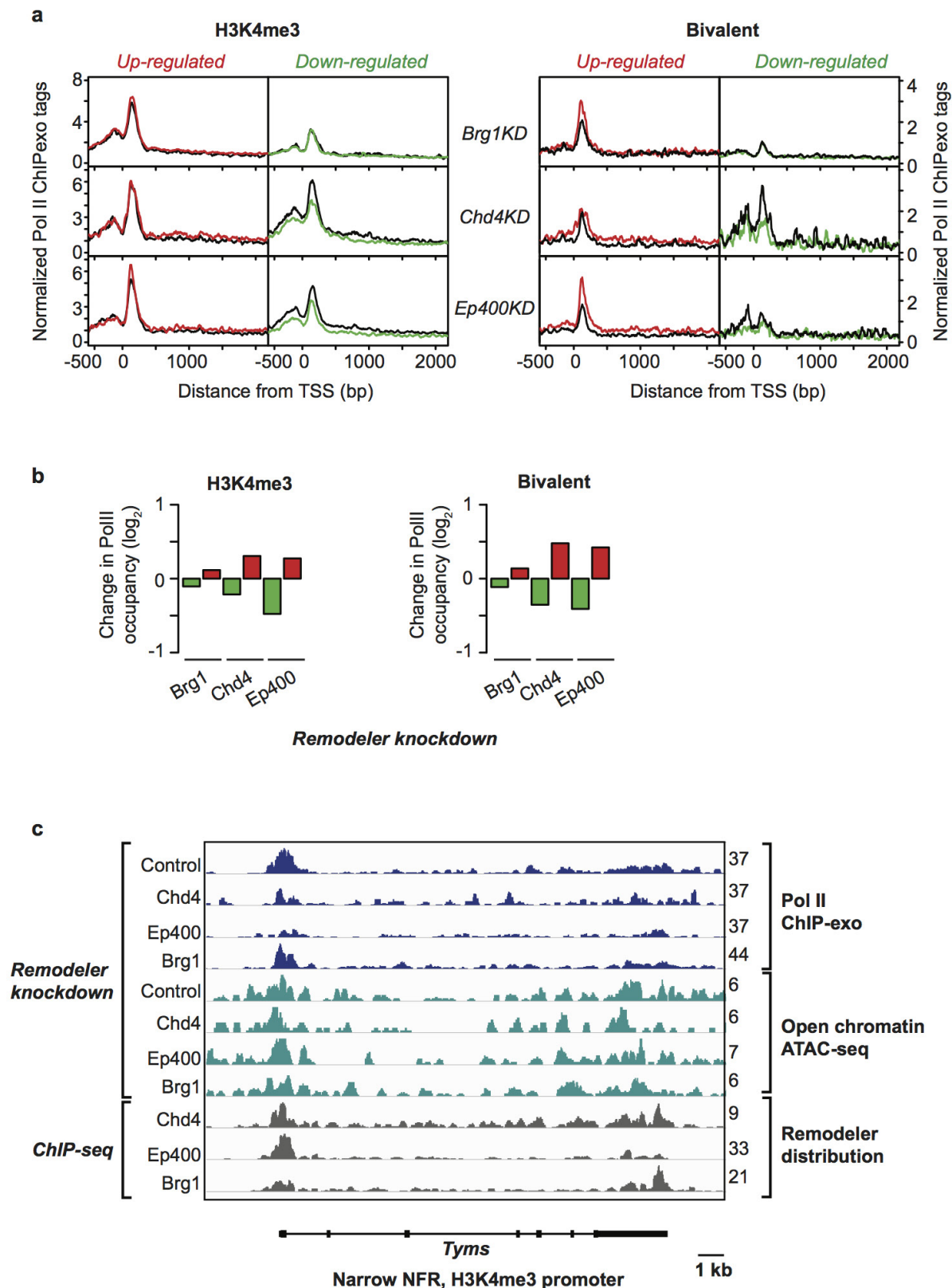
corresponding genes upon remodeler depletion in ES cells. Two distinct shRNA vectors (shRNA 1 and shRNA 2, see Methods) were used for each remodeler. Scores on the y axis indicate the relative expression of the indicated genes compared to reference genes. Values are mean and s.d. of three independent transfection experiments.





**Extended Data Figure 8 | Effect of remodeller depletion on chromatin accessibility at promoters.** Consequence of remodeller depletion by shRNA vectors on ATAC-seq average profiles at all H3K4me3-only (top)

and bivalent (bottom) promoters. Two replicate experiments are shown on each graph for both remodeller knockdown and controls.



**Extended Data Figure 9 | Analysis of Pol II distribution at promoters in remodeler-depleted ES cells. a,** Average Pol II distribution (ChIP-exo) profile in control ES cells (black) or after indicated remodeler knockdown (colour) at H3K4me3-only (left) and bivalent (right) genes. Left and right panels within a set represent the set of genes that are most upregulated (red) or downregulated (green) after remodeler knockdown. Pol II occupancy is indicated within a window spanning 500 and 2,000 bp on the upstream and downstream side of the TSS, respectively. All promoters

are transcribed from left to right. **b,** Bargraphs showing Pol II occupancy change after remodeler knockdown relative to control, measured by ChIP-exo, at genes either downregulated (green) or upregulated (red) after the depletion of the indicated remodeler. **c,** Pol II distribution of remodeler knockdown at a representative locus. Counts indicate reads per 10 million. Pol II loading is markedly reduced at the *Tyms* narrow NFR, H3K4me3 promoter by either Ep400 and Chd4 depletion, suggesting that these two remodelers contribute to Pol II recruitment.

# LOW-COST JOURNALS PIGGYBACK ON ARXIV

*Peer-review platforms built around online preprint repositories spread to astrophysics.*

ILLUSTRATION BY THE PROJECT TWINS



BY ELIZABETH GIBNEY

**A**n astrophysicist has launched a low-cost community peer-review platform that circumvents conventional scientific publishing — and by making its software open source, he is encouraging scientists in other fields to do the same.

The *Open Journal of Astrophysics* works in tandem with manuscripts posted on the preprint server arXiv. Researchers submit their papers from arXiv directly to the journal, which evaluates them by conventional peer review. Accepted versions of the papers are then re-posted to arXiv and assigned a DOI, and the journal publishes links to them.

By piggybacking on, or ‘overlaying’, the arXiv repository, the journal should operate at a fraction of the cost of conventional publishers and will be free for both readers

and authors, says journal founder and editor-in-chief, Peter Coles, an astrophysicist at the University of Sussex in Brighton, UK. He announced on 22 December last year that the journal was open for submissions. It will go live later this month, once its first papers have undergone review.

Development of the software that powers the journal’s peer-review system was led by Arfon Smith, chief scientist at the popular code repository GitHub. Because the software is open source and available at GitHub, Coles hopes that researchers in other fields will adopt the same platform to create their own open journals. “Just cross out ‘astrophysics’ and write ‘condensed matter’ or anything else, and you’ve got your open journal,” he says.

Similar overlay journals already exist in computer science and mathematics; Tim Gowers, a mathematician at the University of

Cambridge, UK, launched one high-profile example, *Discrete Analysis*, last September. But the *Open Journal of Astrophysics* is thought to be the first of its type in physics. Gowers says that he is excited that the platform behind it is open source because “it potentially reduces the costs for others even further”.

## THE OVERLAY MODEL

Coles believes that conventional journals and their associated costs are no longer needed in fields such as astrophysics and cosmology, because most researchers already both submit their work to arXiv and read papers on it.

“The only objection to just putting things on arXiv is that it’s not peer-reviewed, so why not have a community-based effort that provides a peer-review service for the arXiv?” he says — pointing out that academics already carry out peer review for scientific ►

► publishers, usually at no cost.

Coles himself covered the costs of developing the software platform for the journal, amounting to a few thousand pounds, he says. (*Discrete Analysis* licenses different software and is helped by a grant from the University of Cambridge.)

GitHub is covering the costs of hosting the platform, so the only remaining expense is editors' and reviewers' time, which they give voluntarily, says Coles. If the experiment proves successful and the volume of papers balloons, the journal may eventually have to charge authors a handling fee of a few tens of pounds, he adds. (The journal also relies on the continued existence of arXiv, whose running costs amount to less than US\$10 per paper).

The journal does not have the resources to offer services provided by conventional journals, such as heavy editing of papers. Instead, poorly written articles will be rejected and the authors referred to a list of professional copy-editing services, Coles says.

## GAINING TRACTION

Gowers welcomes the new journal; the arXiv-overlay model is much more likely to succeed, he says, if many examples of it can be seen to be working. The journal has amassed an editorial board with high-profile physicists including Pedro Ferreira, a theorist at the University of Oxford, UK, and Andrew Jaffe, a cosmologist at Imperial College London.

But astrophysicists will not necessarily jump to publish in Coles's journal. Ewine van Dishoeck, an astrophysicist at the Leiden Observatory in the Netherlands, says that she, for one, is unlikely to submit her work there. "We have a small number of well established and high-quality journals in astronomy that everyone respects," she says.

Papers in astrophysics are effectively open already, van Dishoeck points out, because anyone can view preprint manuscripts immediately on arXiv, whereas journals in the field make final accepted versions open after a delay — typically 12 months after publication. An issue for researchers can be slow peer-review of papers, she adds, but the *Open Journal of Astrophysics* has yet to prove that it can be faster.

Whatever their costs, the main problems facing all new journals hoping to achieve traction among researchers are ensuring speed and editorial fairness, adds Andrew King, a cosmologist at the University of Leicester, UK. "Reliability — and particularly fairness — are very hard to guarantee," he says, pointing out that the backing of long-lived organizations with a stake in the future of a field, such as learned societies, is often crucial to a journal's success. ■

# Q&A Alice Allen

## The code librarian

*By day, Alice Allen runs software and IT training programmes for the Board of Governors of the Federal Reserve, the US central bank, in Washington DC. But in her spare time, she edits the world's largest registry for software in astrophysics and astronomy research — the Astrophysics Source Code Library (ASCL; ascl.net).*



### How did you come to edit the ASCL?

I've always been interested in astronomy. So in 2010, I started doing volunteer work for the website Astronomy Picture of the Day. I told one of the site's creators — Robert

Nemiroff, an astrophysicist at Michigan Technological University in Houghton, which hosts the ASCL — that I had time for another project. We talked about an effort that he and John Wallin (now at Middle Tennessee State University in Murfreesboro) had started in 1999, to create a repository of astrophysics source codes — the old ASCL. The resource had gathered about 40 source codes, but lay fallow at the time for want of an editor. I took it over and have been working on it ever since.

I work on the ASCL in my spare time and I take vacations to speak about it at conferences. It's an all-volunteer organization: we have two developers and an associate editor, Kimberly DuPrie, a programmer at the Space Telescope Science Institute in Baltimore, Maryland. The advisory committee is made up of astrophysicists who do this because they think it's a good idea. There's a lot of passion around this project.

### Why is the site important?

It increases the discoverability of code used in research. Like many other sciences, astrophysics has become more dependent on software. And as software use has increased, the transparency and reproducibility of the science has decreased — you can read a paper and may not be able to see the source code that enabled the results. The ASCL holds nearly 1,200 records of source codes that have been used in research in peer-reviewed publications. Each entry is citable with a unique ID and points to a website where the code can be downloaded; entries include a description of the code, its authors and some of the research it appears in. We also house some source codes. The site got more than 100,000 hits in 2014. It provides a way for journals to point to a software record — it has been cited more than 500 times since 2012, according to NASA's Astrophysics Data System Abstracts Service.

### How do entries get added to the library?

For the majority of entries, Kim and I look through research papers specifically to find codes to register, and we e-mail the authors; we know that coders won't necessarily think to register their codes with the ASCL. But since a site redesign in 2014, about 40% of the listings have come from scientists submitting their own entries, which we verify. We don't look at the quality of the code, but it has to meet our criteria, such as being used in research and immediately available to download.

If a link goes bad, we track down the code's new home. Journals don't have to worry about references to websites going stale: their links will stay good because they point to the ASCL.

### How is it funded?

It's mostly unfunded. We have in the past received a few thousand dollars to cover things like conference presentations and poster production costs, from organizations such as the American Astronomical Society and the Heidelberg Institute for Theoretical Studies in Germany. Sometimes my travel to conferences is funded, but often it is not. Over the long term, the goal is to make the ASCL financially stable, although a large part of it will always be volunteer-based.

### Do other science disciplines have equivalent code libraries?

There are lots of code registries and repositories online, but they're usually not science-specific. We index codes that are useful specifically in astronomy and astrophysics. There isn't one site for all science codes, and maybe there should be.

I've had several people approach me about creating a similar resource for physics. So last December, we started offering clones of the ASCL infrastructure (which runs using open-source software) to any discipline that wants to build a code registry of its own. A sample site is at scicodes.net, and Michigan Technological University is willing to host other science-code registries for three years if people would like. Nobody has taken us up on this offer yet, but they are welcome to try it out. ■

BY JEFFREY M. PERKEL

This interview has been edited for length and clarity.

CHRIS MADDALONI/NATURE



# CAREERS

**RETIREMENT** With grey hair comes fun dreams for years ahead **p.121**

**PROJECT MANAGEMENT** A job in building scientific collaboration **p.121**

**NATUREJOBS** For the latest career listings and advice **www.naturejobs.com**

AKINDO/GETTY



## COUNSELLING

# Knowledge is power

*A strong network of mentors provides a wealth of experience from which to learn.*

BY PAUL SMAGLIK

Principal investigators (PIs) have a major role in advancing the careers of the junior scientists in their labs. They serve as bosses, role models, teachers and technical advisers. But often, those tasks prevent them from filling another essential need: mentor.

“A PI by definition must be an authority figure,” says Kelly Edwards, associate dean for student and postdoctoral affairs at the University of Washington’s graduate school in Seattle. “Someone who will be tough and strict.” But students and postdocs often need other kinds of support, she says: a counsellor, an advocate, a sympathetic ear, or even just “an aunt who will slip you \$5 every once in a while”.

These sorts of relationships can be hard to cultivate, especially for young scientists whose lives revolve around their labs. But making the effort to step away from the bench to socialize and network at conferences, seminars and university events can pay off later in the form of

valuable career advice and guidance. Mentors can help junior researchers to make difficult career decisions, such as whether to pursue a non-academic career path or switch their research emphasis. And having mentors other than a PI is comparable to having an informal board of directors: the junior scientist benefits from the viewpoint and perspective of multiple people at various career stages.

Some institutions have begun to encourage young scientists to build a mentor network by connecting them with alumni or hosting retreats that put them in direct contact with scientists from different professional backgrounds. These initiatives remove the need for junior researchers to build an entire mentor network on their own. The University of Washington’s graduate school, for example, offers regular sessions to help students tap into the university’s alumni group and other resources. It also requires that students’ individual development plans, — which document young scientists’ training and are required by many

grants — include a section on mentor-team development. Part of the university’s scheme involves telling both the PI and the student that mentors outside the department, or even the university, might be useful for goals such as exposure to industry or soft-skill development.

Similarly, Indiana University–Purdue University Indianapolis hosts regular career-planning workshops in which professionals from the area meet with students, postdocs and faculty members. The university also offers training workshops for young faculty members.

Katherine Shives, a microbiology graduate student at the University of Colorado Anschutz Medical Campus in Aurora, can attest to the usefulness of such events. It took her almost a year and four lab rotations to settle on her PI. But she needed just a few minutes at a university social function to find a key mentor. She started chatting with immunologist Kathryn Holmes, a professor emerita at the university, and felt an immediate rapport. “I felt very comfortable talking to her,” says Shives, who ►

► adds that Holmes has become an important mentor. “I spent a pretty decent amount of time in her office. She’s given me excellent advice.”

Holmes talked Shives through the stress of her first year in graduate school, and has given her the best advice she has received so far in her programme: “Even good change is stressful.” That phrase became Shives’s mantra and has helped her to maintain her perspective during other difficult periods. Holmes also provided guidance as Shives rotated through labs while trying to choose her PI. When the options were carved down, she helped Shives to choose between an established PI with a large lab and a newly fledged PI with a smaller group.

“She advised that established PIs have a solid track record of funding and mentoring students, and that with established labs you know where the PI is and what they are doing,” says Shives. “With younger PIs, there’s no track record of funding and they may have never mentored a student, but there is also the potential of working with a rising star before they get big, and that being the only student means you can get more individual attention.” Ultimately, Shives chose the smaller lab, where she remains today.

She has also stayed in contact with her undergraduate adviser and her PI from an internship during her final year at secondary school, who have also been valuable members of her team, in part because they are not nearby. “Having someone off-campus and removed from my programme was very valuable in the first year,” Shives says. “It helped me keep in perspective that it was not just me struggling, and that the difficulty was not a sign that I was incapable of the work.”

A series of professional interactions through different stages of his education and career helped Tshaka Cunningham to construct his mentoring team. Now a scientific-programme manager at the US Department of Veterans Affairs in Washington DC, he had won a scholarship to attend Princeton University in New Jersey as an undergraduate. Donald Graham, then chairman of *The Washington Post*, was a member of the scholarship committee, and became part of Cunningham’s mentor team after the student requested a personal meeting with him. Molecular biologist Arnold Levine advised Cunningham during his undergraduate programme. After Cunningham graduated, the two stayed in contact; Levine helped him with job interviews, and continued his mentoring role with Cunningham when the younger scientist joined Rockefeller University in New York and Levine became its president. And later, while Cunningham was working as an industry researcher in drug discovery, he connected with a company executive who remains one of his mentors.

Cunningham says that he especially values the support and guidance he received from his mentors when he was conflicted about whether to accept his current job. His quandary was that he loved basic research, but wanted a



**Paul Sauer recommends taking courses with students from other disciplines.**

more-even work-life balance. His mentors counselled him to evaluate what he most wanted, and the advice proved sound — eight years later, he remains pleased with his decision to choose the work-life balance.

He warns, however, that although it is important for junior scientists to feel comfortable with their mentors, they should also look for people who will challenge them. “You want people always in your corner — but who can be brutally honest,” he says. “I seek someone who knows me well, and who has a good enough relationship with me that they don’t worry about hurting my feelings.” His team helped him to weigh up the effort and sacrifice involved in an academic career, he says, and to understand that he could have professional success outside that — even though academia had been his dream.

#### MAINTENANCE WORK

Once a junior scientist has built a network of mentors, the work is not over — he or she must continue to nurture the relationships. Cunningham stays in regular contact with his mentors, but is careful not to smother them. He sends short e-mails a couple of times a year asking how they are doing and what they are working on. Shives meets her mentors regularly for coffee or lunch, and makes a point of catching up with those who are based outside her area whenever they are in town. She makes sure to thank them all for whatever help and support they offer.

Cunningham points out that it is important to remember that a mentor is not a therapist who can devote endless amounts of time to a junior scientist’s challenges. He asks for help only when he has a specific problem. That process helped him to remember both the power

and limitations of mentors. “At the end of the day, a mentor can give you only a framework,” he says.

Mentors outside the lab can help to relieve pressure on the PIs, because the relationships can meet needs that the PI might overlook or be unable to address. This is especially important when a student wants to explore skills or careers away from the bench or outside academia, says Edwards. “That can be a vulnerable point for the faculty member — ‘This trainee doesn’t want to be like me,’” she says. Often, she adds, a senior faculty member simply does not have enough information about non-academic or non-research career pathways to properly guide a junior scientist into or through them.

Mentors need not always be senior researchers: peers can successfully mentor one another.

The European Molecular Biology Laboratory (EMBL) has fostered an initiative in which new students take courses together and learn from each other’s skills. Such a structure can form a strong section of a professional network, says Regis Lengrand, head of administration at EMBL in Grenoble, France.

Paul Sauer, a third-year PhD student studying cellular signalling at EMBL, has gleaned several benefits from the initiative. Taking the courses is an effective and easy way to build peer-mentor relationships — and to set the stage for later collaborations, he says. “Once your own project develops and you are looking for collaborators, or some advice from another field, you can turn to those peers,” he says.

More-senior graduate students and postdocs at EMBL are tasked with teaching workshops, which puts them in touch with scientists at other sites. EMBL also hosts seminars given by external postdocs, companies and visiting faculty members, which provide opportunities for graduate students and postdocs to network and develop external mentor relationships. It also makes its alumni network available to students and postdocs.

The US National Institutes of Health has created a similar model in its Broadening Experience in Scientific Training programme, which sponsors ‘boot camps’ for new graduate students that encourage them to form both peer and faculty mentor networks (see *Nature* 525, 147–148; 2015).

A strong mentor group can give a junior scientist lifelong help with career progression — not just through graduate school and postdocs. “Having a diverse group of mentors, both inside and outside science, was really helpful for me as I discovered my own truth for the career path I wanted to pursue,” says Cunningham.

Shives appreciates the continuing presence in her life of one of her off-campus mentors from early in her education. She still turns to him for emotional support. “You need someone to help you keep your feet on the ground,” she says. ■

**Paul Smaglik** is a freelance writer in Milwaukee, Wisconsin.

SARA SILVA

## COLUMN

## The question I hate the most

Eleftherios P. Diamandis considers when to retire.

About 30 years ago, I emigrated to Canada to pursue my scientific training. For the past 25 years, my laboratory at Mount Sinai Hospital in Toronto has sought biomarkers for use in the early detection of cancer. I love my job as both clinician and scientist. I am now 63 and people throw all kinds of questions at me owing to my diverse experience — and my white hair. But the one I hate the most is: “When are you going to retire?” I hate it because it reminds me that I am in transition. The first time I was asked it, I was 58. The question was unexpected yet it stirred something in my head. And so, I resolved to record whenever people asked me it. At the age of 59, I was asked twice; at 60, four times; at 61, eight times; at 62, sixteen times; and at 63, thirty-two times. By extrapolation, I expect that next year the question will pop up 64 times and by 67, I will be facing it twice a day.

When asked, I pretend not to hear or I whisper something such as “I have no plans to retire” or “I’m still very young”. But when I am alone in my office, the question percolates in my mind. I examine my face in the mirror for fresh wrinkles and ask myself: “Do I really look that old?” The thought that I might be in decline often prompts me to analyse my annual research output for downward trends — papers published, citations received, grants awarded and so on. My favourite statistic is the *h*-index, a measure of both productivity and quality. I love this indicator because it can only rise with age.

Although my lab is still producing good science, I now delegate more tasks, such as grant and manuscript writing, to younger colleagues. And to prepare for my transition into retirement, I have shaped my future plans around my existing passions.

Cyprus, my homeland, is visited by millions of tourists every year. So I thought that I might build a rock ‘n’ roll hall of fame on some land I own there. I could induct

my musical heroes — the Beatles, the Rolling Stones, Pink Floyd, Led Zeppelin, AC/DC — and play their music while collecting entry fees. I might even make enough money to pay the bills. But because this would mean being so far from my grandchildren, I have shelved the idea, for now.

Another plan I made was to host a radio programme in Toronto. I could get paid to play and discuss my favourite music. In fact, a local station did once offer me a daily one-hour show. But my wife intervened by pointing out how such a regular commitment would interfere with our plans for travel and could only lead to stress.

I also considered becoming a full-time grandpa to two girls under the age of five, but when my grandchildren stay with me for just a weekend, I am left completely exhausted. Regular child-care duties are not an optimal occupation for an ageing academic.

So what might be the optimal choice? Staying in the lab. Throughout my career, I was blessed to forge relationships with diagnostic companies and to obtain patents for some of my lab’s discoveries. Because research grants are becoming more difficult to obtain, I created an account that accumulates the resulting royalties and commercial donations. The fund should be large enough to support a couple of graduate students or postdocs, as well as my research manager, for up to ten years without further funding. This represents about 10% of my current lab staff, which I could handle easily as a mentor and adviser. I would be free to pop into my office at any time, to read *Nature* and *Science*, and to write manuscripts or articles like this. I could visit my grandchildren and then return them to their parents.

Nowadays, I consider retirement to be a continuous process that occurs in small increments, over a long period. This slow transition is allowing me to answer my most hated question, as follows: I am retiring at the pace of one minute per day. ■

Eleftherios P. Diamandis is division head of clinical biochemistry at Mount Sinai Hospital, Toronto, Canada.  
e-mail: ediamandis@mtsinai.on.ca

TRADE TALK  
Neuro connector

Dorothy Jones-Davis is a scientific project manager at the Foundation for the National Institutes of Health in Bethesda, Maryland, where she helps companies, non-profit organizations and other institutions to collaborate on projects to understand Alzheimer’s disease.

DOROTHY JONES-DAVIS

### Why did you decide to leave academia?

After completing my PhD, I moved on to a postdoc and was considering faculty positions because that’s what I was surrounded with. What changed my path was that most of the work I found myself doing was considered to be service work — teaching, mentoring, school governance.

### Why does service work interest you?

I seek to connect. I worked to increase diversity in science by exposing kids to science who had not experienced it firsthand. While at the University of California, San Francisco, I sat on high-level committees concerned with diversity in academia, child care and the on-campus culture. I was then awarded a policy fellowship with the aim of broadening participation in engineering. I discovered that I loved influencing the direction of science as well as making connections between folks who should meet. My current job merges those interests.

### How did you prepare for your role?

I did my research before coming to the interview. I had noticed that the foundation lacked programmes in autism or epilepsy, so I pitched ideas to address this, such as bringing together companies that were investing in those fields. My interviewers recognized that I understood what the foundation did. Now I manage a partnership between 31 external organizations for a neuroscience-imaging initiative.

### What advice do you have for job seekers?

There is no cookie-cutter way to achieve your career dreams. In the end, it’s your road and your willingness to do new things. Everyone is searching for the perfect position, but the honest truth is that you make the position wonderful for you too. ■

### INTERVIEW BY MONYA BAKER

This interview has been edited for length and clarity. See [go.nature.com/axdrci](http://go.nature.com/axdrci) for more.



# MUSINGS ON TIME TRAVEL

*An investment for the future.*

BY ROBERT REED

I used to be like him. I was an ignorant, preclucid beast, physically swift but weak in time. Forty years of my life was spent and lost, nothing gained but four exceptionally dull decades. But change became possible and I made the change. I found my ground and transformed my nature, becoming a statue of superchilled glass eagerly racing across the centuries.

The valley stretches before me, green and then black. Green and black.

Up the valley the beast approaches, and reaching me, he stops moving and slows his nature. I don't know how he slows himself. A temporary trick invented in recent years, presumably.

"Let me introduce myself," he roars.

"Slower," I say. "Quieter."

My mind speaks, my hardware lending sounds to the thoughts.

"Sorry," he says.

"Is this better?" he asks.

If I wanted conversation, it would be an improvement. But my hardware says nothing. There's no point being rude, it decides.

Not yet, at least.

Green and black. Green and black. And the beast offers his name, adding, "I respect what you're doing. You are a pioneer."

Pioneers are dead people lost to history. But I am very much not-dead.

"A brave, bold strategy," he says.

I assume that he learned my archaic dialect while preparing for this important, uncomfortable conversation. Every word feels careful. Thought out but always delivered with a self-conscious unease.

"You are practically a legend," he says.

"Practically a legend." That's what people tell you when they're close to forgetting you.

"And I've come here to offer you a service," he says.

Oh shit. A salesman.

I say nothing and the green valley turns black. Night comes and stars wheel, and he says, "Two hundred and fifty-three years crossed so far."

I can measure time as well as he can.

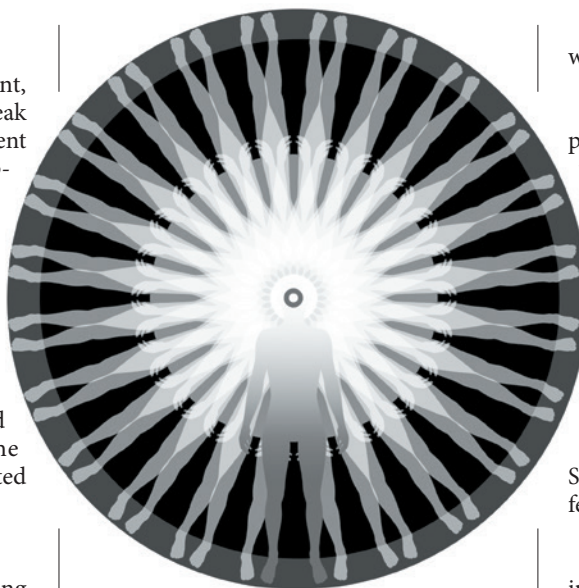
"And you aren't ten days older," he says.

No, I'm three centuries old. I just don't show it.

He smiles and says: "Like you, I think about time travel."

Here we go. The salesman's pitch.

"We're all time travellers," he says. "You're just a lot better at it than me."



Very true.

Then he says, "Sir, do you know why time travel was invented, Sir?"

My hardware could ask him to leave. My AI lawyers, the best in the Solar System, could slap muscular injunctions on him. And to that end, I begin to string together the appropriate commands.

Except then he says: "Consider the first life, the simplest, tiniest cell."

Okay.

"Think what it had to do to survive," he says.

The Sun rises. A cloudy day, with rain rushing in from the ocean on my horizon.

"To survive," he says, "life had to find a complex form. And that form, the good body, had to eat. Grab energy and reproduce. But how does life get enough food? Unless you're a plant, there was never enough nourishment in one physical location. That's why animals grew legs. To graze, to hunt. But of course even a plant has to move. The algae on a rock. There isn't enough sunlight in one nanosecond to feed a little chloroplast. To thrive, the microbe has to journey through time. Which is my point: time travel was invented by our oldest ancestors. Because where life is possible, even if it is only rooted in one spectacular patch of ground, the ignorant entity still has to cross the days and seasons, just on the hope of enduring."

A salesman, yes.

But interesting nonetheless.

"Interesting," is the word I offer.

He agrees with my praise, adding: "But what if?"

I bite. "What if what?"

"If the great goals of life could be accomplished outside time. For instance, what if there was a scenario in which the entirety of your life could play out inside one cold nanosecond?"

"No," I say.

More than once.

"But you shouldn't dismiss me," he says, standing in a sudden rainstorm. "Not if the end result is a vast improvement over your present life. Which is what you did before, isn't it? You didn't enjoy your old life, sluggish and limited. So you remade yourself, accomplishing what few others would even dare attempt."

He knows his market, that's for certain.

Rain finishes with a red Sun plunging into the sea, and the world spins us through space, and I ask: "Why would I want such a life?"

"Because inside that instant, every thought will be possible. Every daydream, every creative urge. Every every everything will be laced into a single eruption of genius. And as I have been assured, that perfection won't last for an instant. Not to the person enjoying it, no. To those special few pioneers, it will seem like a euphoric blaze lasting ten billion years."

I'm not ignorant. My hardware keeps me up to date with human events and the march of technology.

This is news to me.

"Is this even remotely possible?" I ask.

The Sun rises behind me, and squinting into the blaze, he admits: "Not yet, no. The basic research has only just commenced. And even if it were possible, the projected costs would be in excess of what your trust funds could afford."

All becomes clear.

"But it will become possible," he says. "And your investments will eventually make you wealthy enough."

This is a salesman with a very long view.

"Give the idea time," he suggests.

"I will," I promise.

And with that, my new friend walks back down the valley, his long shadow growing short before he vanishes into my thoughts. ■

**Robert Reed** is the author of several hundreds stories and a few novels. He won a Hugo before it was controversial. He lives in Lincoln, Nebraska.

ILLUSTRATION BY JACEY